基于标签本体的MARC元数据扩展机制研究*

□ 魏　来 / 中国科学技术信息研究所　北京　100038
/ 东北师范大学计算机科学与信息技术学院　长春　130117
王雯霞 / 东北师范大学计算机科学与信息技术学院　长春　130117

摘要：MARC元数据应用范围广泛，但其结构复杂，格式设计和思维设计也存在缺陷，通过标签本体与MARC元数据的映射方法能够有效的解决这些问题。标签本体与MARC元数据映射方法的实质是通过建立两者之间的映射对照表，将MARC元数据用标签的RDF三元组形式进行描述，目的是从用户需求的角度提供信息资源的多种检索途径。

关键字：标签本体，MARC元数据，映射
DOI：10.3772/j.issn.1673—2286.2014.02.009

MARC元数据是以代码形式和特定结构记录在计算机存储载体上，用计算机识别与处理的目录，是为描述、存储、交换、处理、检索图书馆书目资源而精密设计的一种元数据格式标准，是目前我国图书馆书目记录描述采用的主要方式。随着网络技术的发展与网络信息资源数量的急剧增长，用户对资源的需求已经不单纯的满足于图书馆所提供的馆内和馆际之间的书目记录及其实体，而是需要更多的关联信息，帮助用户找到更多的资源线索，发现更多的相关资源，这些关联信息可能来自于科学博客、社交网络、学术论坛、维基百科、个人网页等。这些相关网络资源具有一个共同的特征，都能够为用户提供资源标注的功能，即用户使用自己的词语（标签）来描述网络资源，标签是用户对资源内容的理解性描述，通过标签可以揭示用户之间、资源之间的联系，便于从多个角度描述信息资源的内容。因此，通过标签能够在图书馆MARC书目数据与相关网络资源之间建立起一定的关联，实现MARC元数据的有效扩展。

1 标签本体及标签本体的属性特征

标签本体就是利用本体克服标签的语义模糊性、描述方式多样化等缺陷。目前，典型的标签本体模型是Newman 、SCOT 、MOAT 。

Newman标签本体的核心概念是用户(Tagger)、标注(Tagging)、标签(Tag)，其中标签类(Tag)是具体标签(tags)的集合。Newman的标签模型利用FOAF中的agent概念确定用户(Tagger)，利用SKOS中的concept概念确定标签类(Tag)，利用DC中的date概念确定标注行为(Tagging)的时间。Newman标签本体的特点是在标签中实质性的引入本体概念，被其他标签本体广泛利用，并已经在网络上实现普遍应用。Newman标签本体模型如图1所示：

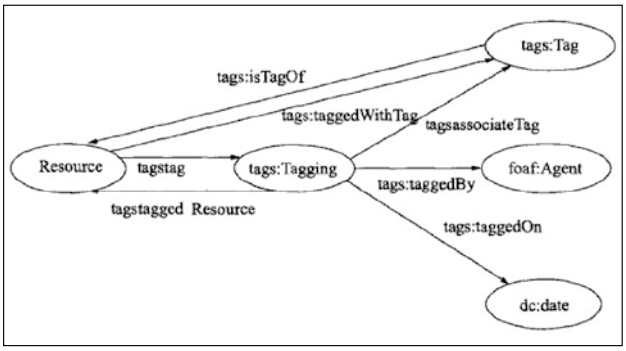

图1 Newman标签本体模型
（资料来源：http://www.holygoat.co.uk/projects/tags/）

SCOT即Social Semantic Cloud of Tags，其核心概念是标签云(TagCloud)、标签(Tag)、共现(Cooccurrence)。在SCOT属性中，scot：contains属

* 本研究为教育部人文社会科学基金青年项目“基于语义化标注的网络学习资源组织方法及实证研究”（编号：11YJCZH180）和中国科学技术信息研究所预研项目“中文社会标注系统语义关联识别方法及实证研究”（编号：YY201202）的研究成果。

性揭示了TagCloud与scot：Tag之间的关系，即Tag归属于TagCloud；scot：hasUsergroup表示使用标签云、参与标注活动的用户群；scot：spellingVariant属性可以连接含义相同的标签；scot：synonym属性可以解决标签含义不明确的问题；scot：frequency属性可以反映标签的利用频率，揭示标签的流行度；scot：taggingActivity属性可以揭示标签云TagCloud和标注行为Tagging之间的关系。总之，SCOT应用广泛，提供了一个跨用户、跨系统共享标签，全面揭示标签语义信息的社会结构。SCOT的模型简化图如图2所示。

MOAT即 Meaning of A Tag， 其定义了标签的全面含义(global meanings)和标签的本地含义(local meanings)。在全面含义(global meanings)中：moat：hasMeaning属性可以把标签本身和标签代表的所有含义连接起来，即可以使Tag和Meaning相连，进而使每个标签都具有全部含义；moat：meaningURI属性可以把每个含义和知识库中所给含义的URI连接起来，即可以使Meaning和Resource URI相连，进而揭示赋予标签这个含义的用户。在本地含义(local meanings)中：RestrictedTagging类可以通过tagMeaning属性把相关标注行为产生的标签连到符合语境含义的Resource URI。MOAT的模型简化图如图3所示。

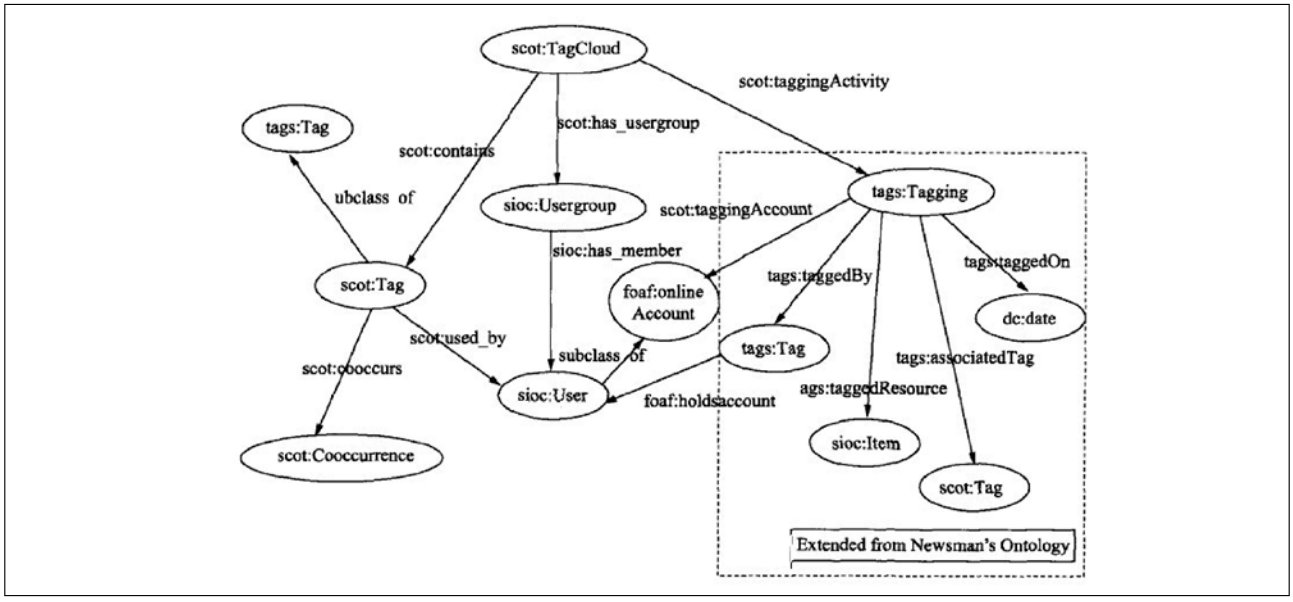

图2 SCOT模型简化图

(资料来源：http://scot-project.net/scot/spec/scot.html)

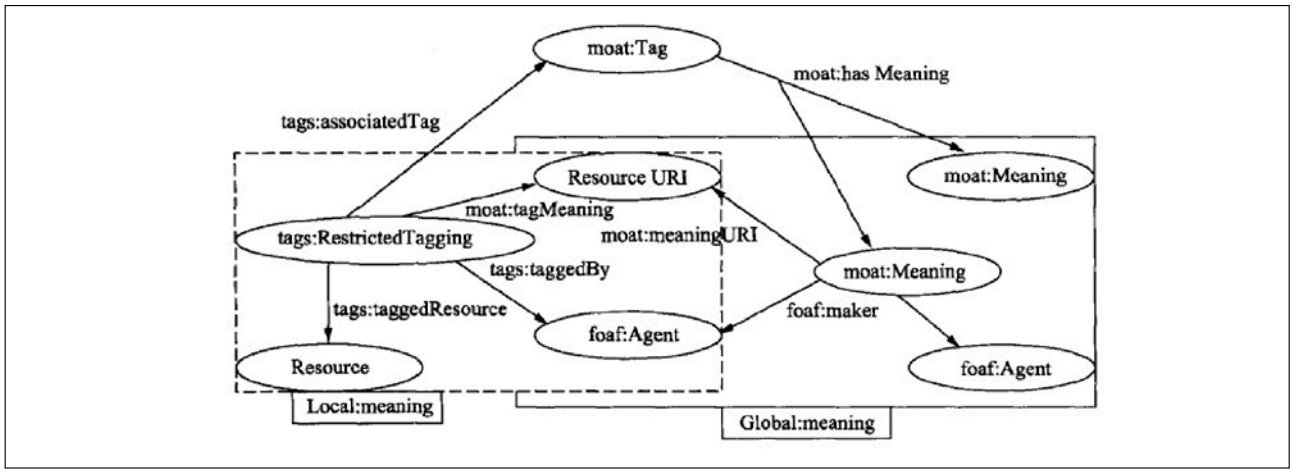

图3 MOAT的模型简化图

(资料来源：http://events.linkeddata.org/ldow2008/papers/22-passant-laublet-meaning-of-a-tag.pdf)

2 基于标签本体的MARC元数据扩展可行性分析

2.1 MARC元数据的局限性

MARC的主要目的是使书目记录形式标准化，以便于利用机器进行规模化处理，但随着对资源组织及其关联与发现需求的提高，传统MARC元数据的扩展性与关联性呈现出了一定的局限性，主要表现在以下几方面：

2.1.1 MARC结构复杂，字段重复

MARC设置了众多字段，优势是尽可能地实现对书目记录的详尽编目，而且MARC复杂的结构要求编目人员在实际运用时要精通MARC中每个位置上字符的不同含义，这容易导致不同人员在理解上的差异，甚至削弱了MARC详尽编目的优势。

2.1.2 MARC的格式设计存在缺陷

MARC最初以磁带为主要存储介质，其编码格式遵循的是磁带格式的国际标准ISO2709，同时MARC的著录、管理和检索都必须利用图书馆的专门软件，限制了MARC资源的开放性和适用性。

2.1.3 MARC的思维设计存在缺陷

MARC的基础是卡片目录，其思维设计也受制于卡片目录，这与用户所需求的数据之间存在矛盾，所以MARC应该从用户需求角度入手，为其提供详尽的数据。1998年，国际图联(IFLA)正式推出FRBR(书目记录的功能需求)报告，FRBR利用实体—属性方法构建了一个揭示书目结构和关系的概念模型 。FRBR根据用户需求将编目对象分成若干层次，这突破了MARC平面化层次的传统思维设计，利用编目对象中的深层次关系建立了一个多层次等级立体的元数据新框架。

2.2 标签本体与MARC元数据映射的可行性

目前，MARC是发展最成熟的元数据格式，是图书馆进行书目资源描述的重要参考依据。标签是用户对资源内容的理解性标注，可以从多角度揭示信息资源的内容，而且标签具有三元组的结构，能够将用户、资源、标签联系起来，更重要的是，标签和MARC之间具有一些共同的属性，有利于两者之间建立映射。两者之间的共同属性包括：第一、两者都包含与数据分离的元素，标签中的格式是属性对应一个属性值，MARC中的格式是字段与内容一一对应。第二、每个元素都有明确的说明，便于准确发现两者之间的相同或者等效之处，例如Newman标签本体中tags:name表示标签含义，MARC中字段200表示资源的正题名。第三、两者对元素的取值都有严格的规范，标签本体就是利用本体技术对标签进行概念规范，而MARC中的取值也有受控词表的限制。由于标签本体与MARC元数据之间这些描述信息资源的共同属性，因此两者之间可以实现映射。

3 基于标签本体的MACR元数据映射

MARC的数据字段区将有关文献的数据按功能块、字段、子字段或数据元素这三个层次组织起来，其中明确加以标识的最小数据单元称为数据元素。在数据字段区的可变长字段内，数据元素构成字段，或若干子字段构成字段 。标签本体的一个重要构成元素是属性，包括对象类型属性和数据类型属性，其中对象类型属性是描述对象之间的关系，数据类型属性是描述时间、日期、数量等非概念的属性。在本文中，主要以普通图书为例，对标签本体与MARC元数据之间的映射进行探索，得出标签属性与MARC元数据之间的对照表 ，具体结果如表1所示。

4 实例分析及结果验证

4.1 实例分析

通过标签本体与MARC元数据的映射对照表，可以将MARC元数据用标签的RDF三元组形式进行描述。给每个书目资源分配一个URL地址，可以利用URL将每个书目资源和其它的书目资源区别开。在本文中以一个简单的实例分析说明将MARC元数据用标签的RDF三元组形式进行映射的方法。选择东北师范大学图书馆的一条MARC机读记录，如图4所示。

表1 标签本体与MARC元数据的映射对照表

标签本体的属性	MARC字段	子字段内容
bibo:isbn	010($a)	ISBN
dc:language	101($a)	正文语种
tags:name dc:title	200($a)	正题名
skos:prefLabel skos:altLabel	200($e)	副题名及其他题名信息
dc:publisher	210($a) 210($c)	出版、发行地 出版者、发行者名称
dc:date	210($d)	出版、发行日期
dc:description	215($a) 215($b) 215($c)	文献数量及单位 载体形态其他细节 文献尺寸
skos:note	300($a)	一般性附注
skos:keyword	606($a) 606($x) 606($y) 606($z)	论题主题 论题主题的论题 论题主题的地名 论题主题的时代
skos:class	690($a)	中国图书馆图书分类法分类号
dc:creator	700($a) 701($a) 710($a) 711($a)	个人名称——主要知识责任 个人名称——等同知识责任 团体名称——主要知识责任 团体名称——等同知识责任
dc:contributor	702($a) 712($a)	个人名称——次要知识责任 团体名称——次要知识责任
dc:source	801($a) 801($a)	记录来源国家 记录来源机构
foaf:agent	905($a)	馆藏信息

图4 MARC机读记录
（资料来源：东北师范大学图书馆书目数据）

图5 MARC记录字段

将MARC元数据用标签的RDF三元组形式进行映射的实质是基于RDF建立起标签属性的数据模型，利用标签本体的各种属性，并使用XML语言进行描述。依据表1即标签本体与MARC元数据的映射对照表，将如图5所示的MARC记录字段建立成基于RDF的标签属性数据模型，其具体的实例图示如图6所示。

图6的RDF句法描述为:

<?xml version="1.0"?>

<rdf:RDF

xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"

xmlns:skos="http://www.w3.org/2004/02/skos/core#"

xmlns:dc="http://purl.org/dc/terms"

xmlns:bibo="http://purl.org/ontology/bibo"

xmlns:foaf="http://xmlns.com/foaf/0.1/">

<rdf:Description

rdf:about="计算机操作系统的URL">

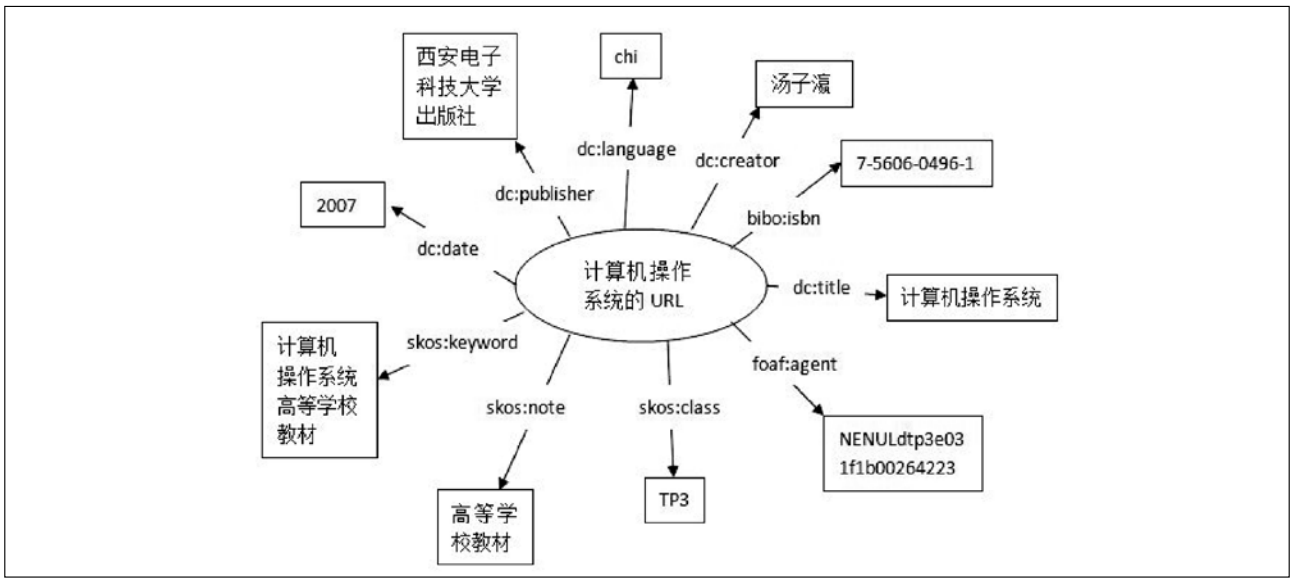

图6 基于RDF的标签属性数据模型的实例图示

<dc:title>计算机操作系统</dc:title>
<dc:creator>汤子瀛</dc:creator>
<bibo:isbn>7-5606-0496-1</bibo:isbn>
<dc:date>2007</dc:date>
<dc:language>chi</dc:language>
<dc:publisher>西安电子科技大学出版社</dc:publisher>
<skos:class>TP3</skos:class>
<skos:keyword>计算机</skos:keyword>
<skos:keyword>操作系统</skos:keyword>
<skos:keyword>高等学校</skos:keyword>
<skos:keyword>教材</skos:keyword>
<skos:Concept>
<skos:note>高等学校教材</skos:note>
</skos:Concept>
<foaf:agent>
<foaf:name>NENULdtp3e031f1b00264223</foaf:name>
</foaf:agent>
</rdf:Description>
</rdf:RDF>

4.2 结果验证

标签是由用户自由创造的，可以真实地反映用户的信息需求，从多角度描述信息资源。而在标签本体与

图7 豆瓣读书标签

MARC元数据之间建立映射的目的是通过标签扩展MARC元数据的检索词汇，向用户推荐更多贴近其信息需求的资源导航，便于用户快速地定位到所需的信息资源。本文中选择如图5所示的MARC记录字段，在豆瓣读书中选择相对应的图书标签，如图7所示。

在如图5所示的MARC记录字段中，从字段606可以看出在MARC格式中对图书《计算机操作系统》选择的关键字是计算机、操作系统、高等学校、教材。在如图7所示的豆瓣读书标签中，可以看出网络用户对图书《计算机操作系统》常用的标签是操作系统、计算机、考研、教材。而通过在标签本体与MARC元数据之间建立映射，可以将标签“考研”增加到MARC对图书《计算机操作系统》的描述中，为用户在检索时提供一个新的途径。

可见，通过标签本体与MARC元数据的映射方法，可以从用户需求角度为其提供更全面、更符合其实际需求的资源导航，便于用户更有效地检索海量资源，从而实现查全率、查准率的最大化。

参考文献

[1] Newman,Richard.Tag ontology design[EB/OL].[2014-01-19].http://www.holygoat.co.uk/projects/tags/

[2] SCOT. [EB/OL].[2014-01-19].http://scot-project.net/scot/spec/scot.html

[3] MOAT. [EB/OL].[2014-01-19].http://events.linkeddata.org/ldow2008/papers/22-passant-laublet-meaning-of-a-tag.pdf

[4] 罗红燕,李章平,陈绍兰. MARC、DC、MODS、FRBR 等文献编目元数据比较[J]. 图书馆学刊，2009(12):25-27

[5] 段明莲. 文献信息资源编目[M]. 北京:北京大学出版社, 2000：113-166

[6] 孙华, 郑巧英. MARC与DC元数据的映像与转换[J]. 上海交通大学学报, 2003, 37:247-249

[7] 黄伟红, 张福炎. 基于XML/RDF的MARC元数据描述技术[J]. 情报学报, 2000, 19(4):326-332

作者简介

魏来（1976-），女，东北师范大学计算机科学与信息技术学院副教授，中国科学技术信息研究所博士后。E-mail: weil875@nenu.edu.cn
王雯霞（1990-），女，东北师范大学计算机科学与信息技术学院2012级情报学硕士研究生。

Study on the Mapping Method between Tag Ontology and MARC Metadata

Wei Lai / Institute of Scientific and Technical Information of China, Beijing, 100038
/ School of Computer Science and Information Technology, Northeast Normal University, Changchun, 130117
Wang Wenxia / School of Computer Science and Information Technology, Northeast Normal University, Changchun, 130117

Abstract: The application range of MARC metadata is very wide and its structure is complex. Moreover, its format design and mode of thinking is flawed. But the mapping method between tag ontology and MARC metadata can effectively solve these problems. The essence of the mapping method between tag ontology and MARC metadata is establish a mapping table and then the MARC metadata could be described by the form of RDF. The aim of the mapping method is to provide a variety of information resources retrieval approach by considering the requirements of the user.

Keywords: tag ontology, MARC metadata, mapping method

（收稿日期：2013-11-29）