

基于实体-关系模型的标准元数据关系研究

赵青青

(上海市质量和标准化研究院)

摘要: 本文梳理了编目学中“关系”的发展历程,结合实体-关系模型,分析标准文献包含的实体、属性,并基于实体间的关联关系和属性关系,开展标准元数据关系研究,同时提出下一步应用方向,包括加强实体关系抽取技术研究、构建标准标签集、开展知识链接研究等。

关键词: 标准元数据, 实体-关系模型, 实体, 关系

DOI编码: 10.3969/j.issn.1674-5698.2023.03.002

Research on Relationships of Standards Literature Metadata Based on Entity-relationship Model

ZHAO Qing-qing

(Shanghai Institute of Quality and Standardization)

Abstract: This paper systematically studies the development of relationship in cataloguing study, and analyzes the entity and attribute of standards literature based on the entity-relationship model, then it conducts the research on standards literature metadata according to association relationship and attribute relationship. Eventually it gives the next research direction, including strengthening the research on entity relationship extraction technology, building standards tag set, and constructing knowledge linking.

Keywords: standards literature metadata, entity-relationship model, entity, relationship

1 引言

在现代汉语词典中,“关系”的第一个基本解释是“事物之间相互作用、相互影响的状态”。“关系”在编目学中一直占有重要作用,编目学中“关系”更多被称为“书目关系”,是界定两个及以上书目实体以某种方式发生关联所产生的特定关系^[1]。

在编目学研究史上,人们对书目关系的研究经历了漫长的过程,在这个过程中产生了很多编目条

例,其中书目关系隐含在各种编目条例中。1980年,国际图联(简称“IFLA”)出版的UNIMARC标准中,首先将MARC元数据的书目关系概括为垂直关系、平行关系、年代关系。在关系发展历程中,西方编目界学者不断对书目关系进行理论扩充,2001年,Barbara B. Tillett女士基于各种编目条例,对书目关系做了更细致与深层次的揭示,提出了7种完整且彼此独立的书目关系,即描述关系、整体/部分关系、附属关系、连续关系、共有特性关系、等同关系、衍

作者简介: 赵青青, 硕士, 工程师, 研究方向为标准馆藏建设与标准信息服务。

生关系^[2],为未来书目关系的建立提供了理论基础。

传统的书目关系多以文献为基本揭示单位,随着用户需求的不断深入,文献资源的知识单元也逐渐朝着细粒度的方向转移,而书目关系揭示方式也必须符合用户的深度需求^[3]。1998年,IFLA正式推出《书目记录的功能需求》(简称“FRBR”),首次拓宽了书目关系的研究范围。FRBR借鉴数据库开发中常用的“实体-关系”模型(简称“E-R模型”),通过描述书目数据中实体和实体间的关系,建立起书目记录之间、书目记录内部各著录字段之间的关系。FRBR打破了传统的书目记录之间彼此独立没有关联的现象,随着FRBR研究及实践的深入发展,世界由此进入基于E-R概念模型进行资源描述与组织的新阶段^[4]。

2 基于实体-关系模型的标准元数据关系分析

2.1 概述

1976年,E-R模型由美籍华人陈品山博士(Peter Chen)提出,E-R模型通过图形化的方式展现数据库设计概念,为数据需求分析提供一条结构化的途径^[5]。在编目界,IFLA基于E-R模型,打破固有思维,重新定位与解析书目关系。本文也将基于E-R概念模型进行标准元数据关系分析。

E-R模型有3个基本元素,即实体、属性、关系。实体是客观存在的具有共同特征的一类事物的集合。实体是可以区分的,可以是人、事、物,或抽象的概念,也可以是属性值;属性是实体的性质和特性,是从不同方向或维度上对实体全方位的揭示和描述,属性的定义以实体为基础,是实体内部与外部特征的提炼和归纳。

关系是实体之间相互影响、相互作用的关联状态。实体关系分为两种,一种是实体间的关联关系,一种是实体的属性。当关系是实体间的关联关系时,其本质是对实体之间存在的显性或隐性关联特性的有效揭示,这种关联关系体现在“实体-关系-实体”三元组模型内;当关系是实体的属性时,其本质是对实体的性质和特性进行的全方位描述和揭示,这种关联关系体现在“实体-属性-属性值”三元组

模型内。上述两种三元组模型是知识网络最基本的构成要素,实体或属性值是知识节点,关系或属性是知识关联。

由E-R模型可知,识别属性和关联实体是元数据关系分析的首要任务。识别属性是通过在特征和特性的描述,使资源能被用户检索到;关联实体则是通过对实体之间外在关系的描述,使资源相互联系,从而帮助用户推荐和提供更多的相关资源。

2.2 识别实体与属性

对于标准元数据而言,实体是用户所关注的关键元数据的描述。用户感兴趣的、可进一步进行知识关联的标准元数据实体包括标准文献、人/机构、标准主题、标准分类、标准术语、适用范围、核心技术内容等。因此,结合标准文献结构特征和标准元数据种类,将进行关系分析的标准元数据实体划分为标准文献、人/机构、主题、分类、术语、适用范围、核心技术内容等七大类,每类实体分别具有各自的属性和属性值。由表1可知,一种实体的属性可以成为单独的实体,例如:标准文献的属性人/机构、主题、分类、术语等可分别成为单独的实体^[6]。标准元数据的关系包括实体与实体之间的关系、实体与实体的属性值之间的关系。

2.3 关联实体

2.3.1 实体与实体之间的关系

根据“实体-关系-实体”三元组模型,此时关系是对实体之间存在的各种显性和隐性关系的揭示。根据表1确定的实体,基于FRBR概念模型书目关系分析方法,实体和实体的关系包括标准文献实体之间的关系、标准文献实体和其它实体之间的关系、其他实体之间的关系。

(1) 标准文献实体之间的关系

标准文献实体之间的关系,包括代替关系、采用关系、引用关系、共有特性关系、衍生关系、等同关系、相关关系。

代替关系:代替关系即年代关系,是标准文献区别于其他科技文献最重要的特性,标准文献的动态更新特点决定了新版标准文献会代替旧版标准文献。代替关系分为完全代替和不完全代替。

采用关系:采用关系是标准文献特有的关系,指

表1 标准文献实体及属性

序号	标准文献	人/机构	主题	分类	术语	适用范围	核心技术内容
1	标准号	起草人	关键词	ICS分类号	术语中文名称	规定内容	章标题
2	标准中文名称	起草单位	主题词	CCS分类号	术语英文名称	适用界限	条标题
3	标准英文名称	提出单位	自由词		术语定义		
4	ICS分类号	归口单位					
5	CCS分类号						
6	发布日期						
7	实施日期						
8	起草人						
9	起草单位						
10	提出单位						
11	归口单位						
12	适用范围						
13	术语						
14	章标题						
15	条标题						
16	主题						

把国际标准或发达国家/地区标准中先进标准的内容不同程度地纳入到新制定的标准中。采用关系包括等同采用、修改采用两类。

引用关系：标准文献的引用关系和其他科技类文献的引用关系类似，即在当前标准中提及或描述另一标准的内容，并列出出处。标准的引用分为规范性引用和资料性引用。规范性引用指引用的文件内容构成了引用它的文件中必不可少的条款；资料性引用指引用的文件内容构成了有助于引用它的文件的理解或使用的附加信息，以参考文献形式列出。

共有特性关系：两个标准文献虽然不直接相关，但拥有共同特性，例如：由相同起草人起草、由相同起草单位起草、由相同归口单位归口管理，拥有相同主题，或者属于相同标准分类等。

衍生关系：当前我国部分国家标准同步发布了外文版，一件标准文献可能同时存在中文版和英文版。同时一件标准文献可能发布修改件、补充件、勘误等，对原标准文本进行修改、补充、勘误等。因此标准文献的翻译版、修改件、补充件、勘误件与原标准文献之间形成衍生关系。

等同关系：随着信息技术的发展和普及，各种类型的数字资源也随之不断增加，标准文献的载体形态除了传统的印本文献外，还包括PDF电子版，未来还可能包括XML格式标准；因此内容相同、载体形态不同的文献之间具有等同关系。

相关关系：除上述关系外，若两个标准文献的

某些特性之间具有上下位关系、相关关系等，则标准文献之间相应形成相关关系。

(2) 标准文献实体和其他实体之间的关系

1) 标准文献实体和人/机构实体之间的责任关系

标准文献实体和人/机构实体之间是责任关系，包括起草人和标准文献之间的起草关系，起草单位和标准文献之间的起草单位关系，提出单位和标准文献之间的提出单位关系，归口单位和标准文献之间的归口单位关系。

2) 标准文献实体和主题、分类、术语等实体之间的主题关系、分类关系、术语关系

标准文献实体和主题、分类、术语等实体之间是主题关系、分类关系、术语关系，主题关系描述标准文献实体的内容，分类关系描述标准文献实体所属的ICS分类和CCS分类，术语关系表明术语来源于标准文献，标准文献涉及相关术语。

3) 标准文献实体和适用范围、核心技术内容等实体之间的整体/部分关系

标准文献实体和适用范围、核心技术内容等实体之间是整体/部分关系，标准文献实体是整体，适用范围、核心技术内容是来源于整体的部分。同样，核心技术内容和其组成部分章标题、条标题、技术内容是整体/部分关系；章标题和条标题是整体/部分关系。

(3) 其他实体之间的关系

1) 人/机构实体的相互关系

标准文献通常由多个起草人、起草单位共同参与起草,这些起草人之间形成合作关系,起草单位之间形成合作关系。由于标准文献未标注起草人和起草单位的对应关系,因此起草人和起草单位之间不存在成员关系。

2) 主题、分类、术语等实体的相互关系,包括同一关系、上下位关系、相关关系

主题、分类、术语等实体的相互关系,可以借鉴叙词表词间关系的类型。叙词表词间关系包括同一关系、上下位关系、相关关系。同一关系指两个概念的外延全部重合,可以相互替代;上下位关系是概念间的等级关系,概括性较强的词语叫做上位词,特定性较强的词语叫做下位词;相关关系是除同一关系、等级关系之外的一种语义相关关系,相关关系复杂多样,包括交叉关系、矛盾关系、对立统一关系型、因果关系、本质与现象关系等^[7]。

若两个标准文献实体的主题、分类、术语等是同一关系,则标准文献实体间属于共有特性关系;若两个标准文献实体的主题、分类、术语等是上下位关系、相关关系,则标准文献实体间属于相关关系。

3) 其他实体的相互关系

人/机构、主题、标准分类、术语、适用范围、核心技术要素等实体之间的交叉关系,在标准文献领域还有待进一步揭示,故此处不予深入分析。

上述两个实体之间的关系都是双向可逆的。

2.3.2 实体与实体属性值之间的关系

根据“实体-属性-属性值”三元组模型,此时属性充当着关系类型的角色。由于每类实体对应的

属性众多,因而每类实体本身及其属性值便可构成关联网络的知识节点,实体和实体的属性值之间的连线即为两者的关系,三者构成一个基本的知识单元。因此属性是知识网络构建过程中的关键节点,是不同于实体之间关联关系的重要纽带。

标准文献是最高层级的实体,根据表1,以标准文献为实体、标准文献的属性为例,标准文献与其属性的关联网络模型如图1所示。



图1 标准文献与属性的关联网络模型

不同标准文献实体之间通过相同的属性值进行关联,形成共有特性关系。例如:多个标准文献具有相同的起草人、起草单位、CCS分类号、ICS分类号、主题词、关键词等,形成同属性关联网络(如图2所示)。

2.4 关系整合

通过整合标准文献实体、属性、关系,得到完整的标准元数据关系表(见表2),其中标准文献元数据涵盖了标准文献的外部特征和内容特征。

3 标准元数据关系应用方向

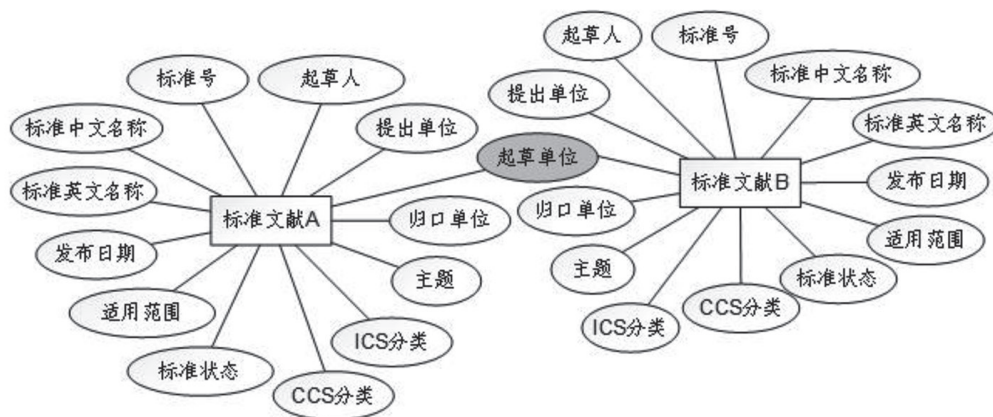


图2 标准文献之间的同属性关联网络模型

表2 标准元数据关系表

实体	关系	实体/属性
标准文献	代替关系、采用关系、引用关系、等同关系、衍生关系、共有特性关系、相关关系	标准文献
	属性关系（标准号）	标准号
	属性关系（标准中文名称）	标准中文名称
	属性关系（标准英文名称）	标准英文名称
	属性关系（ICS分类号）、标准分类关系	ICS分类号
	属性关系（CCS分类号）、标准分类关系	CCS分类号
	属性关系（发布日期）	发布日期
	属性关系（实施日期）	实施日期
	属性关系（起草人）、责任关系（起草人）	起草人
	属性关系（起草单位）、责任关系（起草单位）	起草单位
	属性关系（提出单位）、责任关系（提出单位）	提出单位
	属性关系（归口单位）、责任关系（归口单位）	归口单位
	属性关系（适用范围）、整体/部分关系	适用范围
	术语关系	术语
	属性关系（章标题）、整体/部分关系	章标题
	属性关系（条标题）、整体/部分关系	条标题
	属性关系（主题）、主题关系	主题
适用范围	属性关系（规定内容）	规定内容
	属性关系（适用界限）	适用界限
术语	属性关系（术语定义）	术语定义
章标题	整体/部分关系	条标题
人/机构	同一关系（同起草人、同起草单位、同发布单位、同归口单位）、合作关系	人/机构
主题	同一关系（同主题）、上下位关系（上位主题、下位主题）、相关关系	主题
标准分类	同一关系（同分类）、上下位关系（上位分类、下位分类）、相关关系	标准分类
术语	同一关系（同术语）、上下位关系（上位术语、下位术语）、相关关系	术语

3.1 加强实体关系抽取技术研究，开展标准知识图谱应用

标准元数据关联网络的构建基础是标准文献实体、关联关系的抽取、建立与维护。定义好标准文献实体、属性和关联关系后，需要通过人工或半人工、自动的抽取方法抽取标准文本中所包含的各类实体关系，并将其映射到知识图谱的实体关系表达上^[8]。对于有清晰规律的标准文献实体或属性，比如：标准号、起草单位、采用、代替等，可以采用基于规则的机器标注方法；对于没有清晰规律的标准文献实体，比如：标准中技术指标，通常采用有监督的机器学习，通过邀请领域专家对标准对象进行人工标注，后续开展有监督的机器学习模型训练，并不断优化完善模型。

3.2 开展标准标签集构建研究，促进标准文献知识开发和利用

标准标签集是描述标准全文内容和元数据，分

析标准层次结构和技术要素的工具，通过XML等结构化通用标记语言，定义标准前言、引言、主体等层次以及标准章节条、列项、图、表、公式等技术要素的标签和属性定义，构建通用的标准信息模型^[9]。对于PDF格式的标准文献，通过OCR处理后，提取标准文献的文本内容，将标准文本内容与标准标签集进行匹配关联，可以实现传统PDF电子文档转化为XML格式，进而批量获取相关实体、关系和属性^[10]，实现标准文献的知识关联，为后续进一步开发和利用标准文献奠定基础。

3.3 开展基于关联关系的知识链接构建研究

知识链接是通过知识对象之间特有的属性和关联关系，将离散的知识联系起来的过程^[11]。知识关联是知识链接的基础，在进一步明确不同标准元数据之间的关联关系后，通过建立知识链接，可以构建统一的知识体系结构。标准文献之间通过作废代替关系、引用关系、采用关系等形成联系，也可以通过起草人、

（下转第66页）

6.4 完善技术保障

完善规范合格证技术服务专家队伍和机构,为不具备出厂检验能力的企业开展出厂检验,依托“质量基础一站式”建设工作,为企业打通检验机

构的地域限制,实现当地受理、异地检验、结果互认,大力推进检验机构与企业之间共建检验实验室,共享检测资源,多方位保障产品的出厂检验。

参考文献

- [1] 聂婕,陈琼,王文君,等.美国产品质量监管机制[C].第十五届中国标准化论坛.cnki,2018.
- [2] 王岳平.德国提升产品质量的做法及对我国的启示与借鉴[J].经济研究参考,2012,51:33-37.
- [3] 丁毅,葛健,郭慧馨.日本产品质量监控组织体系及相关制度研究[J].当代经济,2016(8):121-123.
- [4] 刘凌志.我国产品质量监管体制的现状、问题及对策[J].湖南行政学院学报(双月刊),2012,5:42-45.
- [5] 黄星星,刘梦沅.湖南率先推行产品合格证可追溯平台[J].中国质量万里行,2022,1:33.
- [6] 戴铮,李潇,谢小龙,等.基于区块链的产品合格证服务云平台的技术与应用研究[J].现代信息科技,2022,6(2):124-127.
- [7] 武明昆,王小琳,郭志刚,等.一种食用农产品合格证追溯打印技术平台,CN211998094U[P].2020.

(上接第20页)

起草单位等同一性,或者标准分类、主题、术语等上下位关系、相关关系形成相互渗透、相互作用、相互联系的知识集合。揭示了知识的关联关系后,用户可以通过知识链接,最大限度地获取相关知识,从而实现智能检索、知识导航、知识服务等目标。

4 结语

厘清标准元数据的关系是开展标准知识服务的底层支撑。本文的标准元数据是基于标准文献的共性结构提出,并未包括最核心的标准指标元数据,同时标准元数据之间的相关关系还比较宽泛,还需根据具体问题情境,进一步细分优化。下一步,可以借助信息化手段,建立标准元数据的关联,实现元数据关系在标准知识服务中的实践应用。

参考文献

- [1] 罗翀,李菡.解析RDA中的关系描述[J].数字图书馆论坛,2014(06):55-62.
- [2] 高红.书目关系的综合研究[J].图书情报工作,2006(09):108-112.
- [3] 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(04):124-129.
- [4] 成全,许爽.馆藏资源元数据的关联网络结构探析:面向FRBR解构的视角[J].图书情报工作,2014,58(12):124-129.
- [5] 鲁啸,龚葵,魏晨,等.基于E-R模型的情报服务案例库研究[J].情报杂志,2019,38(01):36-40+22.
- [6] 王青.基于元数据的书目关系扩展研究[J].情报杂志,2012,31(09):92-97.
- [7] 刘华.叙词表国际标准的修订及其对基于知识组织的术语服务的影响[J].图书情报工作,2012,56(22):21-25.
- [8] 赵伟,张宽,望俊成.标准文献知识图谱构建的模型设计与集成方法[J].情报工程,2021,7(06):58-66.
- [9] 汪烁,卢铁林,尚羽佳.机器可读标准——标准数字化转型的核心[J].标准科学,2021(S1):6-16.
- [10] 杨跃翔,涂新雨,刘文玲.标准文献知识图谱构建与应用研究[J].数字图书馆论坛,2022(06):22-30.
- [11] 曾建勋.知识链接的构建方式研究[J].图书情报工作,2010,54(12):32-35+77.