

文章编号:1671-8879(2013)06-0079-05

交通信息基础数据元层次结构模型及其应用

张绍阳,高 航,关胜超,曹金山

(长安大学 信息工程学院,陕西 西安 710064)

摘 要:为了改进现有交通信息基础数据元基于业务领域的分类组织造成的冗余问题,提出并建立了交通信息基础数据元层次结构模型。该层次结构模型将交通信息基础数据元划分为基础数据元、抽象数据元及标准规定数据类型 3 个层次。对已经颁布的 JT/T 697 标准进行了层次结构的建立,抽取的 73 个抽象数据元覆盖了 JT/T 697 标准中 55.35% 的交通信息基础数据元。将该结构在交通运输信息数据标准符合性检测中进行了应用。试点工程应用表明:基于层次结构的数据元将标准符合性检测的数据名称的对应率提高了 17.38%;在交通信息基础数据元标准的修订中,采纳了该层次结构模型,有效地规范了数据元的编制工作,该模型在数据元标准的建立和应用方面都具有明显的作用。

关键词:交通工程;交通信息基础数据元;层次结构模型;标准符合性检测

中图分类号:U491;TP39 文献标志码:A

Hierarchical model of basic data element of transportation information and its application

ZHANG Shao-yang, GAO Hang, GUAN Sheng-chao, CAO Jin-shan

(School of Information Engineering, Chang'an University, Xi'an 710064, Shaanxi, China)

Abstract: In order to reduce the redundancy of basic data element of the transportation information which is caused by the classification method in business field, a hierarchical model of the basic data element of transportation information was proposed and built. The hierarchical model was divided into three layers, including basic data element layer, abstract data element layer and standard build-in data type layer. About 73 abstract data elements were extracted from the basic data elements of transportation information, which accounted for 55.35% of the total data elements according to JT/T697 standard promulgated. The hierarchical model was applied to the transportation information data standard compliance testing. Application in pilot project shows that the corresponding rate of data item name with data element name has increased by 17.38% for the use of the hierarchical model in standard compliance testing. In the standard revision of the basic data element of transportation information, the model is adopted and the revision work is standardized effectively. Hierarchical model is significant both in the establishment and application of data element standard. 2 tabs, 3 figs, 6 refs.

Key words: traffic engineering; basic data element of transportation information; hierarchical structure model; standard compliance testing

收稿日期:2013-03-24

基金项目:交通运输部科技项目(2011 364 812 50,2012 364 223 500);河南省交通运输科技计划重点项目(2012D02)

作者简介:张绍阳(1971-),男,山西襄汾人,教授,工学博士,博士后,E-mail:zhshy@chd.edu.cn.

0 引言

交通运输数据具有海量、结构复杂、增长迅速等特点。良好的、一致的交通数据定义对于交通运输数据的有序管理和共享具有重要的作用,《交通信息基础数据元标准》(JTJ 697)的颁布为解决交通信息基础数据元定义的不一致问题提供了标准依据^[1]。

GB/T 18391.2(对应国际标准 ISO/IEC 11179-2)中提出了数据元的 4 种分类模式概念,包括主题词、关键词、分类法与本体论,大多行业的数据元分类都采用分类法进行^[2-3]。在《交通信息基础数据元》中,进一步对分类法进行细化,采用面、线结合的分类方法对数据元进行分类,其中,业务领域采用面分类法,每个类别下设一级、二级和三级分类。这种分类方法业务领域清楚,按照树形结构来组织,便于数据元的查找,但树形结构只是指数据元的分类组织形式,而数据元之间的关系没有进行明确的定义,使得定义重复、数据元的数量大,并且不同的业务领域之间存在相当数量的交叉。例如,在各个业务领域中都规定有人员姓名、单位名称、日期等相似的数据元。张绍阳等提出的数据元二维分类方法综合使用了专业维度和管理维度这 2 种分类方法,水平方向为专业维度分类,垂直方向为管理维度分类,专业维度分类基本沿用交通信息基础数据元的分类方法^[4]。但是该方法仅从数据元的应用角度进行组织,数据元之间的重复仍然存在。林垚对交通科学数据共享工程的数据元进行了分类设计,设置了公路、铁路、水运、民航和其他 5 个大组,并将在应用系统或数据库中有重复、高频率出现、无二义性的数据元集中起来,形成公用数据元,从一定程度上提高了数据元定义的简洁性,但未彻底解决问题^[5]。

为此,本文提出建立交通信息数据元层次结构,使得数据元之间的关系更加明确,在一定程度上消除了数据元之间存在的冗余。

1 交通信息基础数据元层次结构模型

1.1 数据元 3 层结构模型基本概念

交通信息基础数据元定义的核心目的之一就是为交通行业信息系统的建设进行一致的数据类型、格式、单位和值域定义。从这个角度出发,如果将具有相同类型、格式、单位和值域的同类数据元所共同表达的核心概念抽取出来进行统一规定,其同类数据元在定义、归类以及数据元应用时都可以直接引用,则会大大减少数据元的数量,保持数据元的一

致性。基于这样的思想,本文提出了 3 层模型用来建立交通信息基础数据元的层次结构。数据元层次结构模型如图 1 所示。

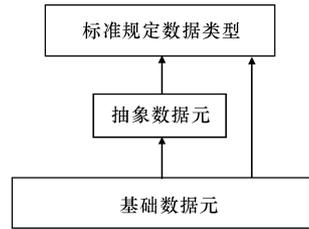


图 1 交通信息基础数据元层次结构模型

Fig. 1 Hierarchical structure model of basic data element of transportation information

第 1 层为标准规定数据类型层。交通信息基础数据元共采用了 6 种类型,包括数字型、布尔型、字符型、日期型、日期时间型、二进制型等。这些类型是以下两层定义的基础。

第 2 层为抽象数据元层。抽象数据元是从交通信息基础数据元中抽取出来的共性数据元,不具有业务领域数据元的明确含义,但是表达了业务领域数据元的特性,故将其称为抽象数据元。例如对 JT/T 697.2 中规定的基础信息数据元“路线起点桩号”、“路段止点桩号”、“圆曲线止点桩号”、“断裂桩号”、“交叉口桩号”、“桥梁中心桩号”等,经分析发现它们具有相同的数据类型、格式、计量单位和值域,并且都是表达一个地址的标识。将上述基础数据元归为一类,提取其共性成分,形成一个抽象数据元——“桩号”,该数据元具有与上述同类的其他桩号数据元相同的类型、格式、计量单位和值域。将“桩号”抽象数据元类型定义为数字型,这样便形成了一个“桩号”类数据元的层次结构模型,如下页图 2 所示。这便是层次结构的朴素思想。下一节从理论上对抽象数据元层的建立进行分析。

第 3 层为基础数据元层。该层的数据元是交通信息基础数据元中规定的数据元,是交通运输业务领域中的不可再分的数据单元。

1.2 抽象数据元层的建立

从上述结构可见,3 层结构建立的核心在于抽象数据元层的建立。下面对数据元名称进行分析并总结抽象数据元的建立方法。

数据元名称中一般包括对象类词、特性词、表示词和限定词^[6]。如:在数据元“安检类型代码”中,“安检”为对象词,“类型”是该数据元的特性词,“代码”是该数据元的表示词。对象词表示数据元所属的事物或概念,它表示某一语境下一个活动或对象,

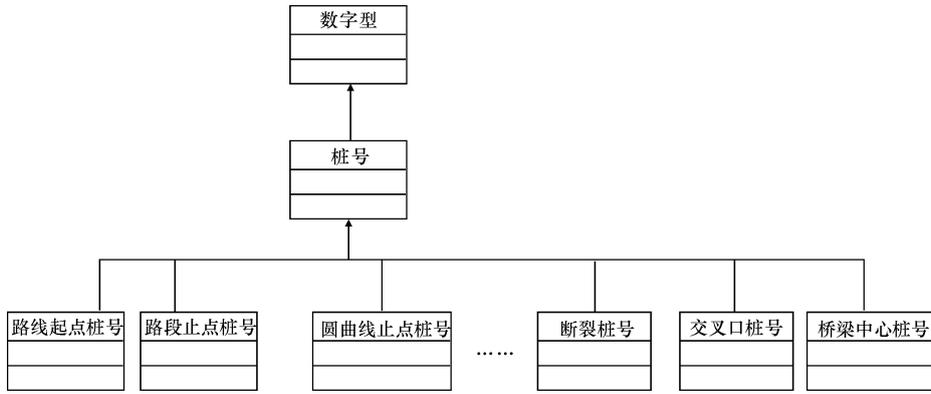


图 2 桩号类数据元的层次结构模型

Fig. 2 Hierarchical structure model of “stake number” data elements

它是数据元中占支配地位的部分,数据元名称中应该有一个且仅有一个对象词;特性词是表示数据元的对象类显著的、有区别的特征,数据元名称中应有一个且仅有一个特性词;表示词是数据元名称中描述数据元表示形成的一个成分,它描述了数据元有效值集合的格式,数据元名称中应有一个且仅有一个表示词。一般情况下,对象词应处于名称的第一(最左)位置。特性词应处于第二位置。表示词应处于最后位置。

为了建立抽象数据元,本文提出以下断言:“在特定的条件下,不同对象的同一特征具有相同的表示形式”,其中,“特定的条件”在此处是指“交通运输领域”以及其他的限制条件。这个断言具有成立的场景,例如“平均行驶速度”、“平均行程速度”等速度类数据元,它们都表达了一个共同的概念,即某对象移动快慢的度量,其中“速”是特性词,“度”是表示词。在交通运输领域,速度都是指宏观物体的速度,例如车辆、飞机、行人等,区别于其他领域例如光学、电学、物理学中的速度,这些物体的速度可以进行统一的表示。又如,前面所举的“桩号”的例子,“桩”是道路设施的一个特性,“号”是其表示形式。对于所有的道路设施来讲,其“桩的号码”都可以统一表示。因此,这种情况下抽象数据元的建立可以概括为:将不同对象类中相同的“特性词+表示词”抽取出来,形成抽象的表示,并对该抽象表示的类型、格式等属性进行定义,则可以形成抽象数据元。

有一些情况不能采用以上规则建立抽象数据元,例如“实验室和研究中心级别和类型是否变更”、“是否经营性教练场”等数据元,它们共同的表示词是“是否”,特性词隐含在其中,直觉上可以形成抽象数据元。但其定义表面上看好像不符合数据元名称定义规则,无法直接利用上述规则。仔细分析可以发现,

这是由于语言习惯造成的,将后一个数据元翻译为标准表达应该是“教练场(对象词)的经营性质(特性词)是否经营性的(表示词)”。对于这些表示词,由于其值域非常明确,无论是何种对象的何种特征,其表示方法都是一致的。因此,这种情况下的抽象数据元可以直接将其“表示词”抽取出来建立抽象数据元。

本文利用上述 2 种规则对抽象数据元层进行建立,当然,与抽象数据元含义一致的其他规则也可应用于抽象数据元层的建立。

1.3 交通信息基础数据元层次结构的建立

依照上述方法,本文对交通信息基础数据元按照层次结构模型进行抽象层次的建立。形成了单位名称信息、时间信息(日期、年度、月份、时间)、联系信息(地址、邮政编码、电话、网址、电子邮箱)、位置信息(桩号、经度、纬度)、度量单位(速度、长度、宽度、高度、车道数、转角、半径、超高值、加宽值、金额、流量等)、人员信息(姓名、性别、年龄、出生日期、职称、职务、学历等)等抽象数据元。根据基础数据元的定义对抽象数据元的类型、格式、值域、单位等进行定义,最终形成交通信息基础数据元的层次结构模型。

下页表 1、表 2 是基于 3 层结构抽象出来的部分抽象数据元及其部分属性。

交通信息基础数据元标准共分为 13 个业务领域,分别由不同的单位进行编制,在 2013 年的修订工作中,《总则》的编制组采纳了本文提出的层次结构的思路,规定了部分常见的抽象数据元,对于各业务领域数据元的修订工作起到了良好的规范作用。

在交通运输部科技项目“交通信息数据与标准规范符合性检测关键技术及规范”研究中,项目组按照层次结构模型对交通信息基础数据元进行抽取,共抽取 73 项抽象数据元,所代表的交通信息基础数据元达 4 000 多项,占数据元总数的 55.35%。

表1 联系信息抽象数据元

Tab. 1 Abstract data of contact information

抽象数据元名称	定义	类型	格式	备注
地址	人员或机构的通讯地址或住址	字符型	an..100	
邮政编码	人员或机构的邮政编码	字符型	n6	
电话	人员或机构的办公电话、传真等各类固定电话	字符型	an..18	
网址	个人或机构的网址	字符型	an..50	
机构名称	机构的法定名称	字符型	an..100	
电子信箱	电子邮箱名称	字符型	an..50	

表2 时间信息抽象数据元

Tab. 2 Abstract data of time information

抽象数据元名称	定义	类型	格式	备注
日期	事件发生的日期	日期型	YYYYMMDD	
年度	日期的年度	字符型	n4	
时间	事件发生的时间	日期时间型	YYYYMMDDhhmm	

2 数据元层次结构在交通信息数据标准符合性检测中的应用

交通信息数据标准符合性检测是指根据国家、交通运输行业信息化相关数据标准(主要指数据元标准)及项目设计文件,对交通运输信息化建设项目中软件系统的数据字典、数据库结构、数据库内容等与标准的符合程度进行检测的活动,主要检测内容包括数据的格式、类型、单位和值域等。通过检测,能为用户提供标准化建议,从而提高交通运输行业信息化的标准化水平。

从检测内容可知,要进行检测,首要任务就是将用户的数据定义与标准的数据元利用其标识属性进行对应。目前,由于数据元的主要标识属性就是数据元名称,因此,将用户的数据项名称与数据元标准名称的对应是标准符合性检测的基础和关键。例如,在数据元标准中规定了“路线起点桩号”数据元的格式、类型、单位、值域等,用户的数据字典中定义了“路线起始桩号”字段,要检测该用户的定义是否遵守了数据元标准,需要对用户给出的数据名称和标准中该数据元名称进行对应,才能利用该标准的规定对用户数据的其他方面进行检测。在项目研究中,采用了编辑距离、增加语境的改进方法、基于语义的对应方法等多种方法提高对应率。在本文中提出了一种基于数据元层次结构的数据名称对应方法。下面对其进行简要介绍。

将用户数据项名称和标准数据元名称对应时可能出现如下情况:

(1)在标准库中找到了用户数据项应该遵守的标准数据元。即在名称对应时,运用其他名称对应技术可以较为准确地将被检测的用户数据名称与标准中某一个基础数据元对应。这时,就可利用该标准数据元对该数据项进行检测。

(2)在标准库中找不到相应的标准数据元。即在名称对应时,其他名称对应技术不能准确地从标准库中找到相应的基础数据元。这种情况下,可以利用抽象数据元对其进行检测。①对于大多数情况,抽象数据元名称都在数据元名称的最后,例如“桩号”抽象数据元,因此,可将抽象数据元名称与用户数据项名称的末位进行匹配,如果匹配了,则表明该数据项可以使用该抽象数据元检测。例如,用户数据项名称为“引线开始桩号”,在标准中未对该数据项进行规定,这时候便可使用“桩号”抽象数据元对其进行判断。②对于抽象数据元不在数据名称最后的情况,例如“是否”抽象数据元,需要在该抽象数据元的记录中设置标记,在比较时,在用户数据项的名称中进行搜索抽象数据元名称,如果找到,可利用该抽象数据元对该数据项进行检测。对于上述2种情况都不能找到抽象数据元的数据项,可认为该数据项未采标。

在《交通信息数据与标准规范符合性检测系统》中,采用了抽象数据元的数据项名称对应方法。检测实例如下页图3所示。“样例数据库”中“日期”、“地址”、“时间”、“周转量”等数据项采用抽象数据元进行检测。其他数据项采用基础数据元直接进行检测,“船舶定期许可证”为未采标数据项。

在试点工程中,对交通运输部“公路水路建设市场诚信及工程质量发布系统”中的某个子系统的数据库进行了标准符合性检测,该系统共使用了1 007个数据项,初步剔除掉编制不规范的和不完整的,有效数据项为863项。使用系统提供的语义对应方法,能够对其中的562项数据进行检测,检测率为65.12%。进一步使用本文提出的基于抽象数据元的数据名称对应方法,系统能够对其中的712项数据项进行检测,检测率为82.50%,检测率提高了17.38%,进一步提高了检测系统的检测能力。通过检测,系统为这些数据项推荐了标准的格式、类型、单位和制约,从而大幅提高了被检系统的标准符合性。



图 3 交通信息数据标准符合性检测实例

Fig. 3 Examples of standard compliance testing of transportation information

3 结 语

(1)在分析现有交通信息数据元及其分类方法中存在问题的基础上,提出了交通信息基础数据元层次结构模型,对已经颁布的 JT/T 697 标准进行了层次结构的建立,抽取的 73 个抽象数据元覆盖了 JT/T 697 标准中 55.35% 的交通信息基础数据元。

(2)将该结构在交通运输信息数据标准符合性检测中进行了应用,在试点工程应用使得标准符合性检测的数据名称和数据元名称的对应率提高了 17.38%。

(3)在进一步的研究工作中,将考虑交通行业信息化的需求建立抽象数据元,而不局限于仅从现有的交通信息基础数据元中抽取。

参考文献:

References:

[1] JT/T 697.1-2007, 交通信息基础数据元第 1 部分: 总则[S].
JT/T 697.1-2007, Basic data element of transportation information part 1: general principles [S].

[2] GB/T 18391.2-2003, 信息技术数据元的规范与标准

化第 2 部分: 数据元的分类[S].

GB/T 18391.2-2003, Information technology-specification and standardization of data elements part 2: classification for data elements [S].

[3] ISO/IEC 11179-2-2005, Information technology-metadata registries (MDR) part 2: classification[S].

[4] 张绍阳,王选仓. 公路信息基础数据元二维分类及其应用[J]. 武汉理工大学学报: 交通科学与工程版, 2007, 31(5): 815-818.
ZHANG Shao-yang, WANG Xuan-cang. 2D-classifying method of highway information foundation data element and its application [J]. Journal of Wuhan University of Technology: Transportation Science & Engineering, 2007, 31(5): 815-818.

[5] 林 垚. 基于数据元技术的交通科学数据目录设计[J]. 科学技术与工程, 2011, 11(13): 3085-3089.
LIN Yao. Traffic scientific data design based on data elements technology[J]. Science Technology and Engineering, 2011, 11(13): 3085-3089.

[6] GB/T 18391.5-2001, 信息技术数据元的规范与标准化第 5 部分: 数据元的命名和标识原则[S].
GB/T 18391.5-2001, Information technology-specification and standardization of data elements part 5: naming and identification principles for data elements[S].