

基于语义形式化的XBRL链接库优化

张颖敏(博士)

(广州民航职业技术学院, 广州 510403)

【摘要】XBRL链接库用于表示XBRL财务元素之间的关系,具有可选性。本文探讨添加不同链接库形式化表示后的系统推理的效率问题,并通过对个别利润表实例证明该优化方法是有效的,能降低形式化表达的规则数,提高逻辑推理的效率。

【关键词】XBRL; XBRL链接库; 元数据; 语义形式化

一、引言

XBRL(eXtensible Business Reporting Language,可扩展商业报告语言)是公认的企业网络财务报告最有效的形式。XBRL分类标准中含有大量的元数据,用于描述商业事实的背景参数,具有明确的语义。XBRL分类标准采用XML Schema和XLink技术规定这种财务元数据的语义。由于XLink语义表达和推理能力上的欠缺,可能导致因财务概念认知歧义和实际业务差异而误用、误增概念及其约束关系。因而如何运用本体方法,实现财务元数据的语义形式化和一致性控制是产学界面临的主要挑战。

XBRL是基于XML格式的,本身具有XML的良好结构。García等提出了将XBRL转化为RDF格式,以实现形式化表达和数据的连接,但是没有对现行分类标准中链接库的关系表达进行优化。王东提出了XBRL的语义元模型,并应用描述逻辑来对其进行形式化表达,但该研究只是针对单个XBRL的财务数据语义表达,而且没有考虑XBRL的时态特征,也没有对XBRL财务元素语义表达的链接库进行优化。本文将对XBRL元模型中元类的关系表达进行图形化表达,并以时态描述逻辑TDL_{BR}对XBRL的财务元数据进行形式化表示。在此基础上,分析XBRL分类标准中五大链接库关系描述的重复性、必要性和替代性,对链接库的关系表达进行优化。最后通过我国财政部发布的通用分类标准中的利润表元数据进行优化前后的形式化表示比较,以验证该优化方法的有效性。

二、XBRL财务元数据的形式化

由于XBRL分类标准的可扩展性,XBRL财务元数据具有明显的分布式异构的特点,为了加强XBRL元数据的可比性,更好地表示XBRL技术规范的元数据语义,XBRL国际组织于2011年发布了XBRL抽象模型1.0,2012年修订后发布了最新版的XBRL抽象元模型2.0,以概念及建模方式展示了XBRL技术规范中元模型的抽象

表达,可以方便技术和非技术人员对于XBRL的认识,以便开发出相应的应用和接口程序。下面将在抽象元模型的基础上,研究XBRL财务元数据的元类以及元数据的语义的图形表达方法,并对其进行基于TDL_{BR}的形式化表示。

元模型是关于模型的模型,用于描述特定领域中构成模型的元数据结构和语义,为了解决产品数据的一致性和企业信息的共享,国际对象管理组织(OMG)提出了元建模的四层体系结构MOF(Meta Model Facility),该结构元建模的抽象分层体系结构,根据模型的抽象程度从低到高分为4个部分: M₀信息层,是现实世界的对象数据层; M₁模型层,由元数据组成,用于描述数据以及数据的结构等信息,是信息层的抽象,信息则是具体模型的实例; M₂元模型层,是模型层的抽象,用于描述模型层中的元数据结构和语义; M₃元一元模型层,是元模型层的抽象,用于描述元一元数据结构和语义。具体如图1所示。

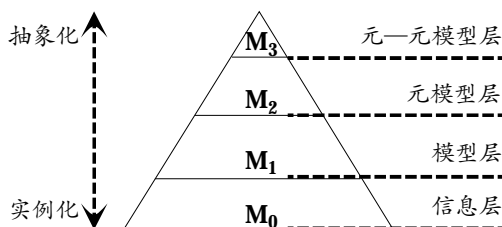


图1 MOF元模型体系结构

在XBRL体系结构中,技术规范作为纲领性文件,用于规定分类标准和实例文档的结构及其语义;分类标准文件描述了XBRL财务数据的定义及其关系,实例文档则是企业汇报商业事实的元素,XBRL分类标准和实例文档组合构成了技术规范的实例;实例文档中商业事实的

具体取值则组成了企业的XBRL财务信息。由此可见, XBRL元数据是元模型的实例化,而XBRL抽象模型则展示了XBRL分类标准概念模型与代表商业事实的实例文档之间的关系,具体如图2所示。

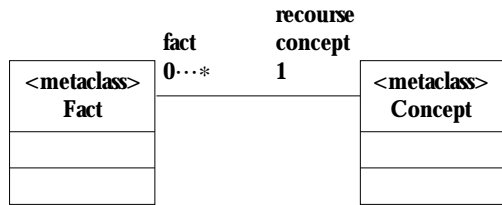


图2 XBRL概念与事实的元类关系

上述XBRL模型结构表明,在XBRL系统中,通过对概念模型中概念以及概念关系的实例化来表示企业中的商业事实,因此,模型中的Concept是XBRL体系中分类标准的抽象,而Fact则是XBRL实例文档中商业事实的抽象。图2展示了Fact与Concept之间的n:1关联关系,其中,Fact是商业事实的元类,Concept则为概念模型的元类,表示一个商业事实对应于一个的概念资源元类,而1个概念资源元类可以与多个商业事实关联,被多个商业事实引用。

描述逻辑的知识体系中,TDL_{BR}的知识库由术语集TBox和实例集ABox组成,根据元数据体系的四层架构, XBRL元模型对应于术语集TBox,可将其形式化为TBox中的概念及关系的定义,而将作为元模型实例的元数据形式化为ABox。

在XBRL模型结构中,元数据的语义表达是在元模型的基础上延伸而来的,是元模型中元类的实例,因此XBRL在描述逻辑TDL_{BR}中的知识表示对应地分为XBRL元模型中的元类以及元类的实例化元数据的形式化两部分。其中,元类的形式化包括对元类定义的形式化以及元类之间关系的形式化,通过对XBRL抽象模型的分析, XBRL元模型中元类间的关系包括继承、聚合和一般关联关系。

XBRL中的元数据是元模型中相应元类的实例,元数据之间的关系则为元类之间相应关联的实例,因此可以把XBRL元数据转化为TDL_{BR}的实例集ABox。相应的, XBRL元数据的形式化分为以下四种情况:

- (1)若元数据中的元素c是元模型中元类C的实例,则可形式化为:C(c)或c:C,本文将采用C(c)形式。
- (2)元数据的继承关系可以看做是元类子集的实例,因此也是元类的实例,即可用上述第一种方式来形式化。
- (3)若元数据中的元素c₁与c₂存在聚合关系,相应的元类C₁通过聚合关系R_{Ag}与元类C₂相关联,则可形式化为:R_{Ag}(<c₁,c₂>)。
- (4)若元数据中的元素c₁与c₂存在关联关系,相应的

元类C₁通过关联关系与元类C₂相关联,而这种关联关系可被形式化为概念R_{As}和角色r₁,r₂,则c₁与c₂之间的关联关系可形式化为:R_{As}(a),r₁(<a,c₁>),r₂(<a,c₂>)。

三、XBRL元素关系的形式化表达优化

XBRL的体系结构中分类标准是表达XBRL财务报告语义的核心,而XBRL分类标准由XML模式文件以及链接库文件组成。其中,模式文件以元素定义的形式出现并定义了对其他可发现分类集模式文件的引入关系,在模式文件中,概念被赋予一个具体的名字和一种类型,该类型根据概念定义来界定事实标准中允许的数据类型。

链接库是扩展的链接的集合。XBRL分类标准中通过五大扩展链接库来进一步表达概念内部的关系以及概念和文档之间的关联关系,而这些扩展链接是可选的,在形式化时选择的链接库文件越多,所能表示的关系越多, XBRL财务报告的语义也越丰富。然而,链接库中基本元素关系的描述数据量非常大,如在我国财政部颁布的XBRL通用分类标准中,核心模式文件描述的基本元素包括2845个,链接库中对这些元素关系的描述也呈线性递增,若对所有的关系都通过描述逻辑形式化出来,其推理规则数也相当庞大,而链接库主要采取XLink技术来表示元素间的关系,过多的链接跳转可能会导致推理的不可判定性。因此,应该选择合适的链接库进行形式化表示,在能保证逻辑语义表达力的同时,也要考虑复杂度,保证推理的可判定性。

可以通过链接库描述的关系入手,分析关系的描述是否为表达财务报告事实所必需的。在此,我们区分XBRL链接库的关系为必要关系和可替代关系两种。具体的步骤如下:首先,分析XBRL财务报告分类标准核心元素在财务报表中所表示的关系;其次,分析哪些关系是必要的,哪些在形式化表示中是可代替的;最后,选择相应的链接库进行形式化表示,并判断其推理规则数的优化程度,对比分析链接库选取前后的推理效率。

1. XBRL的扩展链接关系。XBRL分类标准中主要由五种扩展链接库来说明概念,分别是:定义链接库、计算链接库、列报链接库、标签链接库以及参考链接库。

(1)定义链接库:用于描述元素间不同类型的层次结构以及为元素定义新的链接关系,主要包括以下四种基本的关系类型:

- 第一种,一般—特殊关系,表示一般与特殊的种属关系,是指一个元素项是另一个元素项的抽象,或者一个元素项涵盖了另一元素项,如:资产与流动资产。
- 第二种,原名—别名关系,是指两个元素项在本质上指的是同一个概念,如:资产与实收资本。
- 第三种,必须—元素关系,或跟随出现关系,是指一个元素项的出现依赖于另一个元素项,这种关系主要用于对信息披露的监管。如,当某企业在资产负债表中填写

了生物资产的金额,那么就必须填写生物资产的有关附注。没有附注内容的报告在验证时将不予通过。这就需要在定义链接库中定义生物资产元素与生物资产附注元素的“必须—元素关系”。

第四种,相似—元组关系,是指两套项目组合在结构上相似,用于描述概念间的相似关系。

上述定义关系中,一般—特殊关系代表了财务元素之间的整体和部分的继承、泛化关系,是定义财务元素之间关系的关键,因此,根据上面的划分原则,这是形式化表达的必然关系。

(2)计算链接库:用于定义模式文件中各元素之间的线性计算关系。这是反映财务事实的核心,因此也被划分为必要关系。

(3)列报链接库:用于储存模式文件中元素间的信息,构建商业报表中各元素之间的列报层级关系,使计算机软件能将代码以财务报告列报的形式展示出来。这是生成财务报表的列报依据,因此,也被认为是必要关系。

(4)标签链接库:用于对模式文件中的元素定义个性化的标签,以支持不同语言的业务数据呈现。

(5)参考链接库:用于描述元素对权威文献的参考关系,帮助了解元素的具体含义。这是财务元素含义的权威说明,是财务元数据语义形式化表达的必然关系。

上述的五种扩展链接库中,定义、计算、列报三种扩展链接类型用于管理分类元素之间的关系,标签、参考扩展链接则表达了概念和文档注释间的关系,用于描述财务元素的具体概念定义。

2. XBRL 扩展链接库的优化。XBRL 链接库是可选的,如何选择合适的链接库,既能保证表达力,又能提高推理效率,是 XBRL 财务元数据形式化需要研究的课题。

下面将从 XBRL 链接库中对财务元素关系的描述展开分析,研究关系的必要性和可替代性,在此基础上进行链接库的优化选取。

XBRL 通用分类标准中,对元素的关系描述通过链接库文件中“定位器”(Locator)的 `xlink:href` 属性来指定元素的来源,而元素与元素之间的关系则通过“弧”(Arc)的 `xlink:from`、`xlink:to` 属性来指定,其中 `from` 属性指向的是整体的元素,`to` 属性指向的是部分的元素,而整体部分的关系在列报链接库。具体的关系元类如图 3 所示:

图 3 中,Concept 元类与 Locator 元类是一般关联关系,Concept 端基数为 1,Locator 端基数为 0..*,表示存在 1 个 Concept 对应 0 个或以上的 Locator;Locator 元类与 Arc 元类是一般关联关系,Locator 端基数为 1..2,Arc 端基数为 1..*,表示 1 个 Arc 对应 1 到 2 个 Locator。

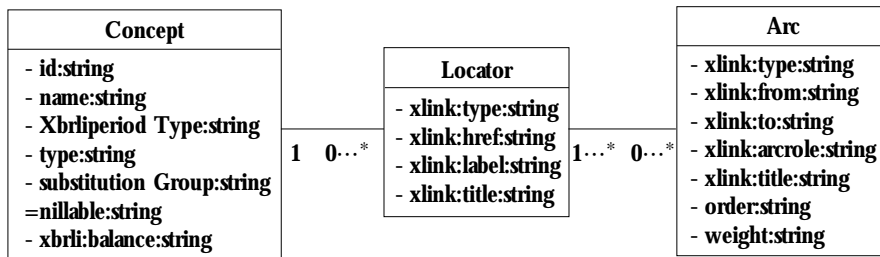


图 3 XBRL 分类标准中基本元素的关系元类

XBRL 分类标准中基本元素是固定的,都来自于核心模式文件,关系的描述主要在五个链接库中。在三大财务报表的定义、展示和计算链接库的关系描述中,可以发现对元素的描述有重复,下面以个别利润表中的营业利润 (OperatingProfits) 与营业收入 (OperatingRevenue)、营业成本 (OperatingCost) 等元素在链接库中的描述为例进行说明,该元素在财务报告中的关系如图 4 所示,即营业利润=营业收入- 营业成本- 营业税金及附加- …- 资产减值损失+ 公允价值变动收益+ 投资收益(详见个别利润表)。其中,元素左边的符号代表加/减关系,右边代表元素在利润表列报中的层级顺序。

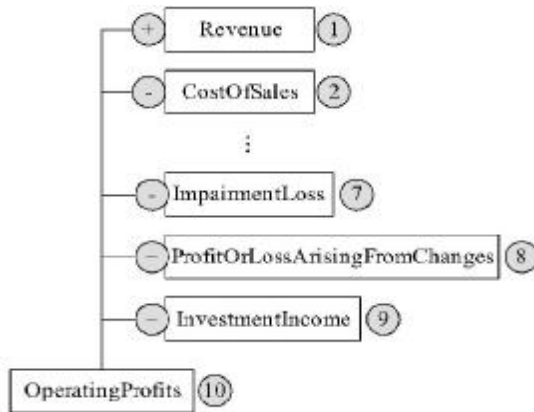


图 4 营业利润与营业收入等元素的关系

以下是营业利润与营业收入的 XML 属性在链接库中的描述片段:

```
<link:loc xlink:title="Operating Profits" xlink:href="..
/./cas_core_2010- 09- 30.xsd#cas_OperatingProfits" xlink:
type="locator" xlink:label="OperatingProfits"/>
```

```
<link:loc xlink:title="Revenue" xlink:href="http://xbrl.
iasb.org/taxonomy/2010- 04- 30/ifrs- cor_2010- 04- 30.
xsd#ifrs_Revenue" xlink:type="locator" xlink:label="Reve-
nue"/>
```

其中,定义链接库、列报链接库与计算链接库对 loca- tor 的描述相同。

类似分析得到营业利润与营业收入元素在定义和列报链接库除了弧的名称以及描述的关系类型不一样外,

其他属性如 **xlink:from**、**xlink:to**、**order** 的值都是相同的,可以选取其中之一来形式化表示,而另一个链接库可通过映射机制获取其表达。此外,由于列报链接库可以表示报表的列报层级关系,而其中的 **parent-child** 关系可以转化为定义链接库的整体一部分、一般—特殊关系,因此列报链接库应为必选项。

计算链接库主要描述的是元素之间的线性计算关系,通过 **weight** 的属性来指定元素间的加(1.0)减(-1.0)的关系,这是表示财务事实的关键,也是其他链接库所不能转化或替代的。因此,计算链接库也是必选的。

综上所述,在进行 **XBRL** 财务元数据形式化表示时,考虑语义的表达能力以及系统推理的可判定性、效率性,本文选取定义链接库的一般—特殊关系、列报链接库层级列报关系和计算链接库来描述 **XBRL** 财务报表标记元数据的关系,而每个链接库中都通过定位器 **Locator** 链接元素的来源,来描述元素的概念,因此,对概念的形式化可只选择链接库其中之一的定位器属性,而关系的形式化则应选择定义链接库的一般—特殊关系,以及列报链接库和计算链接库的弧 **Arc**。

3. 链接库选取后的 XBRL 推理效率分析。描述逻辑是一阶谓词逻辑的一个可判定的子集,描述逻辑的推理任务主要是判断概念的可满足性。对于一个 **TDL_{BR}** 概念 **C**,其可满足性是判断是否能找到一个解释 $I=(V^I, V^T, g^I, g^T)$,使得解释空间 C^I 不为空,即 $C^I \neq \emptyset$ 。如果能找到则说明在解释 **I** 下概念 **C** 是可满足的,否则概念 **C** 是不可满足的。描述逻辑的可满足性、等价性和不相交的推理可以转化成包含关系的描述,因此描述逻辑的一致性问题可以规约到概念的包含关系的判断。**Ding Pan, et al(2012)** 对 **TDL_{BR}** 的 **Tableau** 算法通过构造算子扩展的树的深度越大,则算法的复杂度也随之增加。对于 **XBRL** 分类标准,作为 **TBox** 的集合,其术语的元素个数直接影响到推理的效率。而通过对链接库的选取优化可以减少重复描述的概念定义,替代某些关系的描述,从而大大减少需要形式化的概念的关系数,提高推理效率。

四、XBRL 财务元数据的 TDL_{BR} 形式化验证

为验证上述 **XBRL** 财务元数据形式化方法以及链接库选取优化的有效性,下面选择能反映企业经营状况的利润表进行形式化研究,具体选择我国财政部发布的通用分类标准中财务报表列报中个别利润表(参见右表)。

我国财政部发布的通用分类标准中,对个别利润表的元素的描述分布在核心模式文件 **cas_core_2010-09-30.xsd** 和五大链接库文件中。其中,核心模式文件和标签链接库文件是通用分类标准公用的,而其他链接库文件则放在专门用于描述财务报表的 **cas_30_2010-09-30** 文件夹中。下面将根据 **XBRL** 元数据形式化理论,对上述利润表分类标准进行形式化表示。

个别利润表的部分元素及其关系

元素名称	列报层级 Order	计算关系 weight
利润表 (IncomeStatementLineItems)		
一、营业收入 (OperatingRevenue)	1	1.0
减:营业成本 (OperatingCost)	2	- 1.0
营业税金及附加 (BusinessTaxAndSurcharge)	3	- 1.0
销售费用 (DistributionCosts)	4	- 1.0
管理费用 (AdministrativeExpenses)	5	- 1.0
财务费用 (FinanceCosts)	6	- 1.0
资产减值损失 (ImpairmentLoss)	7	- 1.0
加:公允价值变动净收益 (ProfitOrLossArisingFromChangesInFairValue)	8	1.0
投资净收益 (InvestmentIncome)	9	1.0
二、营业利润 (OperatingProfits)	10	1.0
.....		

1. 财务元素的形式化。如前所述, **XBRL** 元素的定义通过 **Locator** 元素来实现,而 **Locator** 通过 **Concept** 元素来实现对模式文件中基本元素的属性定义,根据 **TDL_{BR}** 的形式化原理,对利润表的元数据形式化如下:

(1) 模式文件的元数据形式化。利润表列报事项元素 (**IncomeStatementLineItems**),是维度化表格的列报事项,这是抽象元素,该类元素在财务报告中没有实际内容,只是为了满足列报效果需要,但对展示元素间的层级列报关系有重要意义,其形式化如下:

```

Concept( IncomeStatementLineItems ),
id( cas_IncomeStatementLineItems ),
type( xbrli:stringItemType ),
xbrli:periodType( duration ),
substitutionGroup( xbrli:item ),
nillable( true ),
abstract( true );
营业收入( OperatingRevenue ):
Concept( OperatingRevenue ),
id( cas_OperatingRevenue ),
type( xbrli:monetaryItemType ),
xbrli:balance( credit ),
xbrli:periodType( duration ),
substitutionGroup( xbrli:item ),
nillable( true ),
abstract( false );
    
```

……(其他个别利润表中的元素基本属性形式化同理,在此不赘述。)

(2) 模式文件中元素概念及其属性的聚合关系形式化。利润表列报事项元素的属性聚合关系如下:

```

RAg( Concept- id ) ( <IncomeStatementLineItems, cas_In-
    
```

comeStatementLineItems>)

$R_{Ag}(\text{Concept-type})(\langle \text{IncomeStatementLineItems}, \text{xbrli:stringItemType} \rangle)$

$R_{Ag}(\text{Concept-xbrli:periodType})(\langle \text{IncomeStatementLineItems}, \text{duration} \rangle)$

$R_{Ag}(\text{Concept-substitutionGroup})(\langle \text{IncomeStatementLineItems}, \text{xbrli:item} \rangle)$

$R_{Ag}(\text{Concept-nillable})(\langle \text{IncomeStatementLineItems}, \text{true} \rangle)$

$R_{Ag}(\text{Concept-abstract})(\langle \text{IncomeStatementLineItems}, \text{true} \rangle)$

……(其他个别利润表中的元素属性的聚合关系形式化同理。)

2. 财务元素之间关系的形式化。财务元素的关系主要包括列报的层级关系和计算关系,下面针对链接库文件中对利润表元素的这两种关系的描述进行形式化表示。

(1)层级关系。财务元素的层级关系可以通过父子类的继承关系来实现,表示父类和子类之间的上下级关系,因此对利润表财务元素的层级关系形式化可以通过继承关系来实现,具体如下:

$\text{OperatingRevenue} \subseteq \text{IncomeStatementLineItems}$

$\text{OperatingCost} \subseteq \text{IncomeStatementLineItems}$

$\text{BusinessTaxAndSurcharge} \subseteq \text{IncomeStatementLineItems}$

$\text{DistributionCosts} \subseteq \text{IncomeStatementLineItems}$

$\text{AdministrativeExpenses} \subseteq \text{IncomeStatementLineItems}$

$\text{FinanceCosts} \subseteq \text{IncomeStatementLineItems}$

$\text{ImpairmentLoss} \subseteq \text{IncomeStatementLineItems}$

$\text{ProfitOrLossArisingFromChangesInFairValue} \subseteq \text{IncomeStatementLineItems}$

$\text{InvestmentIncome} \subseteq \text{IncomeStatementLineItems}$

$\text{OperatingProfits} \subseteq \text{IncomeStatementLineItems}$

$\text{OperatingRevenue} \cup \text{OperatingCost} \cup \text{BusinessTaxAndSurcharge} \cup \text{DistributionCosts} \cup \text{AdministrativeExpenses} \cup \text{FinanceCosts} \cup \text{ImpairmentLoss} \cup \text{ProfitOrLossArisingFromChangesInFairValue} \cup \text{InvestmentIncome} \cup \text{OperatingProfits} \subseteq \text{IncomeStatementLineItems}$

上面的形式化表达式表明利润表中部分元素与利润表列报事项元素之间的继承关系,可以展示利润表中元素的层级关系,最后一个式子表示:利润表列报事项由这些元素共同组成。

(2)计算关系。XBRL分类标准中计算链接库表达了财务元素之间的线性加减计算关系,对于复杂的计算关系可以通过公式链接库来实现,而本文关注的利润表只涉及加减计算关系,因此,只考虑计算链接库的形式化。财务报表中的元素归纳为借—贷关系,为了表示元素之间的计算关系,可以通过借贷相等的关系来获得,即属于借关系的元素集合与属于贷关系的元素集合。

这里引入符号 R_c 、 R_d 分别表示借、贷关系(属性),则

$R_c \cdot C_1$ 、 $R_d \cdot C_2$ 分别表示具有借、贷属性的概念 C_1 和 C_2 。上述利润表元素中,展示了营业收入与营业成本等元素之间的线性计算关系,根据计算链接库中对元素计算权重 **weight** 的设置,其关系可表达为前所述的:营业利润=营业收入-营业成本-营业税金及附加-销售费用-管理费用-财务费用-资产减值损失+公允价值变动收益+投资收益,整理得:营业利润=(营业收入+公允价值变动收益+投资收益)-(营业税金及附加+销售费用+管理费用+财务费用+资产减值损失),因此,该借贷关系可以形式化表示如下:

$R_c \cdot \text{OperatingProfits} = (R_c \cdot \text{OperatingRevenue} \cup R_c \cdot \text{ProfitOrLossArisingFromChangesInFairValue} \cup R_c \cdot \text{InvestmentIncome}) \cap (R_d \cdot \text{OperatingCost} \cup R_d \cdot \text{BusinessTaxAndSurcharge} \cup R_d \cdot \text{DistributionCosts} \cup R_d \cdot \text{AdministrativeExpenses} \cup R_d \cdot \text{FinanceCosts} \cup R_d \cdot \text{ImpairmentLoss})$

根据上述形式化的方法,笔者统计上述个别利润表的 11 个元素中,核心文件、定义链接、列报链接、计算链接、参考链接以及中英文两个标签链接对该元素的描述共包括 110 个,关系个数为 59,对应的形式化表达式个数应为 458,而经过链接库选取优化后的形式化表达式个数为 198,减少了 260 条重复描述,而随着财务元素的增加,其优化效果更明显。

五、结论

本文根据 XBRL 技术规范对元数据的描述,探讨将其图形化的方法,并以时态描述逻辑 TDL_{BR} 为形式化工具对其元类、元数据及其关系进行形式化表示。接着分析分类标准中对关系描述链接库文件的语义表达,发现链接库文件的关系描述中有严重的重复性。针对描述的重复性以及 XBRL 财务报告报送的需求,区分链接库关系表达的的必要性和可替代性,对链接库文件中财务元数据的形式化规则进行精简与优化。最后通过我国财政部发布的通用分类标准 CAS 中个别利润表中表示元素概念的定义、元素属性以及元素之间的一般—特殊、层级、计算关系进行基于 TDL_{BR} 的形式化表示,结果表明经过优化后的处理,能大大降低推理的规则数,提高推理效率。

主要参考文献

Roberto García, Rosa Gil. Triplifying and linking XBRL financial data [C]. Proceedings of the 6th International Conference on Semantic Systems, 2010(3).

王东. XBRL 财务报告元数据语义形式化与推理研究 [D]. 暨南大学博士学位论文, 2013.

S. Staab and R. Studer. Handbook on Ontologies, second edition. International Handbooks on Information Systems [M]. Springer Verlag, 2009.

【基金项目】国家自然科学基金项目(项目编号: 71171097)