

doi: 10.12052/gdutxb.170029

# 基于领域本体的网络财务报告文本信息抽取研究

梁倬骞<sup>1,3</sup>, 王东<sup>2</sup>, 朱慧<sup>2</sup>, 潘定<sup>1</sup>

(1. 暨南大学 管理学院, 广东 广州 510632; 2. 广州大学 工商管理学院, 广东 广州 510006;  
3. 暨南大学 信息学院, 广东 广州 510632)

**摘要:** 企业财务报告中存在大量蕴含着许多重要财务信息的非结构化文本信息, 这类信息难以被计算机识别、分析和处理, 也难以通过数据库技术进行管理. 本文结合本体相关理论和自然语言处理(Natural Language Processing, NLP)技术, 从词语属性描述、词语关系组织和相关知识链接3个维度构建财务报告领域本体, 利用NLP工具对中文财务报告中的文本信息进行处理, 将非结构化文本信息转化为结构化信息并使用XBRL表示, 在一定程度上实现了文本信息的数据库存储与计算机分析处理.

**关键词:** 可扩展商业报告语言; 领域本体; 财务报告

中图分类号: TP391

文献标志码: A

文章编号: 1007-7162(2017)03-0089-07

## A Research on Text Information Extraction from Annual Report Based on Domain Ontology

Liang Zhuo-qian<sup>1,3</sup>, Wang Dong<sup>2</sup>, Zhu Hui<sup>2</sup>, Pan Ding<sup>1</sup>

(1. School of Management, Jinan University, Guangzhou 510632, China; 2. School of Business Administration, Guangzhou 510006, China; 3. School of Information, Jinan University, Guangzhou 510632, China)

**Abstract:** Significant financial information can be retrieved from the vast amount of textual data provided in Chinese business accounting reports (annual reports). Nevertheless, due to the unstructured nature, this textual information usually is difficult to be obtained and analyzed via traditional computer and database techniques. To address this issue, a set of unified domain-specific ontology is presented, combined with Chinese Natural language processing (NLP), which transforms accounting reports in unstructured text into a structured XBRL-based form via three different dimensions, namely word attribute description, word relation organization, and related knowledge links respectively.

**Key words:** extensible business reporting language(XBRL); domain ontology; financial report

可扩展商业报告语言(extensible business reporting language, XBRL)作为目前公认的财务信息交换标准和技术, 通过标记元数据将传统的企业财务报告转化为计算机可以直接识别的数据格式, 提高了数据处理效率. 学界对XBRL的应用研究主要在如下方面: 数据互操作、分类标准及其映射规则、基于本体的语义推理、信息披露质量以及软件工具应用等<sup>[1-4]</sup>. 这些研究集中在XBRL可识别的结构化数据基础上, 然而, 财务报告中包含大量不能被XBRL识

别的非结构化文本块, 它们中蕴含着有关企业的一些重要信息. 这些非结构化形式的数据无法被机器识别, 导致XBRL实例文档中大量文本信息不能自动读取, 因此不能实现计算机自动分析、推理工作<sup>[5]</sup>. 在信息系统领域, 本体已经成为精确语义表达、知识共享和重用的主流方法. 从术语学视角看, 领域本体是关于领域知识的术语及其关系的术语体系<sup>[6]</sup>. 那么, XBRL领域本体就是广泛共享、形式化的财务报告术语体系与相关实例, 也称为形式本体. 它可看作

收稿日期: 2017-02-17

基金项目: 国家自然科学基金资助项目(71171097, 71671048); 中央高校基本科研业务费专项资金资助项目(15JNLH005); 广东省自然科学基金资助项目(2015A030310506)

作者简介: 梁倬骞(1984-), 男, 实验师, 博士研究生, 主要研究方向为信息管理与信息系统.

通信作者: 王东(1984-), 男, 讲师, 博士, 主要研究方向为信息管理与信息系统, E-mail: wangdong@gzhu.edu.cn

是电子化的会计准则,遵循此准则的多种分类标准共享其术语体系.依托XBRL领域本体可导出所需的分类标准,可支持对财务数据的推理校验,从而为自动生成XBRL财务报告铺平道路.

本文旨在解决非结构化的财务文本信息的计算机识别和处理,在构建XBRL财务报告领域本体的基础上,将涉及财务信息的文本逐层解构,通过主题词表、已存本体、结构化词典识别文本中的术语、进行语义关联关系挖掘,将蕴含于文本信息中的数据有机地关联起来,减除资料的异构性.通过术语识别与抽取,提高财务报告中文本内容抽取、分析与处理能力.

## 1 相关研究概述

### 1.1 本体方法在XBRL财务报告中的应用

近年来,会计界开始引入本体理念试图提升分类标准的质量.分类标准被认为是采用XBRL语言把专业认知规则映射成可应用的知识体系,这种映射需要本体理论作为指导<sup>[7]</sup>,而且分类标准工程成为软件工程、知识工程和本体工程的交叉融合学科<sup>[8]</sup>.从技术角度看,2006年以来Lara、Bao等学者认为<sup>[9-11]</sup>虽然制定分类标准时达成了会计概念表达的共识,但大多还只是隐式语义或非形式语义,为支持更广泛地利用XBRL数据,应该将分类标准知识体系转换成领域本体,并解决XBRL与OWL两种语言间不同表示的映射问题.进一步看,多位研究者认为,由于会计处理方法的多样性特点和概念的变化,不可避免地带来不同分类标准中概念语义的差异,需要借助本体实现多种分类标准的语义融合<sup>[12]</sup>.

学术界已经形成利用本体方法增强XBRL语义表达的共识,研究偏重说明本体中应包含的构成要素和增强互操作性,尝试将分类标准直接转换为OWL本体,但并未讨论利用时态描述逻辑构建财务报告术语的精确语义表达,目前描述逻辑对整体-部分关系的描述也存在表达能力不足.

### 1.2 财务报告领域本体的构建

领域本体构建就是模型化一个论域.通常在本体构建中,首先由领域专家人工确定概念、关系和规则集合,然后人工借助本体工具将其形式化. Romilla等提出基于本体的XBRL映射和决策框架(OFXD Ontology-based Framework for XBRL-mapping and Decision-making)<sup>[13]</sup>,该框架通过XBRL财务报表信息的提取自动生成XBRL本体;

Li提出建立XBRL元模型的方法,通过元模型转化为本体并增加规则从而实现语义推理<sup>[1]</sup>;Roberto等运用XML语义重用方法,将XBRL的核心XML文档结构与OWL进行映射,形成XBRL的本体架构;García等运用RDF数据立方体词汇将XBRL与开放数据进行融合,转换成RDF三元组,从关联数据角度实现XBRL与其他类型数据的交互<sup>[14-15]</sup>;吴忠生等以中文财务报告转换为背景,提出一种统一的基于XBRL多分类标准的财务报告本体,并构建基于该领域本体的XBRL财务报告转换模型,促进了不同分类标准之间的财务报告信息转换<sup>[16]</sup>.以上研究集中在XBRL语义扩展方法上,未进一步对XBRL不能表示的文本信息处理进行研究.

### 1.3 财务报告本文信息挖掘

Antonina等用语言分析方法针电信行业领先企业财务报告中信息、文体、语言及可读性进行研究<sup>[17]</sup>,通过实例验证得出企业财务报告和网上公布报告的行文风格相似度差异较大;Ryo ISHIZAKI等学者将财务报告中用文本描述的项目进行了归类,并用自然语言处理方法进行文本挖掘和处理<sup>[18]</sup>.上述研究从自然语言处理方法对财务报告的文本挖掘做出了初步探索,但侧重于词为单位进行挖掘和聚类处理,未考虑文本信息与财务报表信息的映射关系,不能有效体现文本信息与财务报表信息的互动.

基于上述研究,本文针对财务报告中文文本中的财务信息进行领域本体构建,提出以XBRL本体为顶层框架来识别文本信息中的主题词的方法,进而建立主题词与财务报表、主题词与主题词的映射关系,对整块文本中相关财务信息进行抽取,利用XBRL技术将其描述为计算机可读数据进行推理.

## 2 财务信息领域本体构建

### 2.1 财务信息领域本体的构建原则

为了指导和评估本体的设计过程,Gruber等人1995年提出本体设计需遵循“准确、一致、可扩展、编码偏好程度最小和最小本体承诺五项原则”<sup>[19]</sup>,在此基础上结合财务报告领域特点,本文认为财务报告的领域本体构建过程需遵循的基本原则有:

(1) 标准化.参照XBRL技术规范和会计准则通用分类标准,财务信息中对概念和关系的定义必须是标准化的,能够准确一致地描述财务报告领域内的信息.

(2) 唯一性.需保证领域本体描述的每个概念和

关系在全局范围内都是唯一的。

(3) 可扩展性. 财务报告本身随着信息技术以及行业的发展会有调整变化, 财务报告领域本体的模型框架则需具备较好的可扩展性, 支持新的概念或关系添加到已有领域本体中。

## 2.2 财务信息领域本体的构建方法

本文通过建立财务领域知识概念—概念间关系—实例文件之间的语义关系, 实现处理层次从语法层面向语义层面的转换. 首先, 从分类标准层中抽取元素概念及层级关系, 采用OWL语义技术, 建立财务报告领域本体库; 其次, 建立分类标准与领域本体之间的映射, 完成分类标准及其元素语义标记, 抽取多层次分类标准的树形结构, 形成映射规则库; 再次, 按照《企业会计准则通用分类标准指南》规定, 标记实例文件与分类标准结构层级的映射关系, 产生数据转换方案, 生成XBRL报告实例。

## 2.3 财务信息领域本体的构建流程

本文构建领域本体采用自顶向下和自底向上相结合的半自动化方法, 构建流程如图1所示。

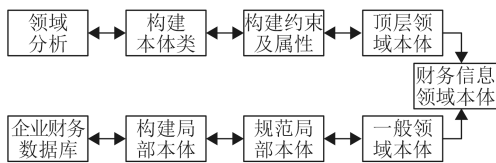


图1 财务信息领域本体构建流程

Fig.1 Constructing process of financial information domain ontology

本文以某保险公司年报中关联方与关联交易模块的一段文本为例, 说明财务信息本体构建、关系描述、文本抽取与XBRL表达的全过程. 文本具体内容如“本公司与集团公司自2003年9月30日以来持续签订有保险业务代理协议. 本公司与集团公司于2008年12月30日签订2008年续展确认书, 将保险业务代理协议续展3年, 有效期自2009年1月1日起生效, 至2011年12月31日止. 根据该协议, 本公司同意向集团公司就非转移保单提供多项保单管理服务. 本公司根据该协议作为服务提供商, 但不享有或承担非转移保单项下的保险人的权利和义务. 保险业务代理服务计费方式参见财务报告“重大关联方关系及关联交易”附注. 在截至2011年12月31日的3个年度, 该交易金额年度上限均为人民币14.02亿元. 本公司与集团公司于2011年12月15日签订2011年续展确认书, 将上述保险业务代理协议进一步续展至2014年12月31日. 在截至2014年12月31日

的3个年度, 该交易金额年度上限均为人民币11.88亿元. 本公司于截至2011年12月31日止年度向集团公司收取保单代理服务费用共计人民币11.12亿元。”

## 2.4 财务信息领域本体构建

### 2.4.1 顶层领域本体构建

顶层领域本体对领域本体中基本的、常识性的概念进行定义. 在XBRL财务报告领域, 分类标准是在可扩展标记语言规范下, 对财务报告信息的概念、结构、关系和限定的系统描述. 其中, 分类标准采用XBRL Schema定义概念, 如通用分类标准下以元素为单位界定报告项目、项目类型、项目实例文件等信息; 采用XLink语法描述概念存在的关系, 如财务信息中包含的层级关系、合并和剥离关系、求和与比例关系等. 因此, 本文采用分类标准为基础构建财务信息的顶层本体, 从分类标准内在层级结构提取领域本体的隐含结构。

具体而言, 该层包括权威部门颁布的财务报告分类标准和企业自定义分类标准. 在披露XBRL财务报告时, 企业首先根据监管要求选择相应的分类标准. 当企业披露的信息并不能全部由该分类标准表达时, 就会自定义财务信息元素来满足披露要求, 从而生成企业自定义分类标准. 尽管自定义分类标准不是由权威部门颁布, 但同样反映了企业的披露特征, 因此同属于分类标准层。

财务报告中文本中蕴含的财务信息是对企业财务情况的重要补充, 以上述文本为例, 其内容既对应于财务报表的主要科目如营业收入、应收账款、银行存款, 又对财务报表附注中关联交易列表的重点内容进行了说明. 因此, 对财务信息的本体构建顶层术语应参照XBRL分类标准, 将文本信息中的主题词以扩展分类标准的形式处理. 根据XBRL分类标准, 借鉴吴忠生等XBRL财务报告转换的方法, 以顶层领域本体构建为目标, 将财务信息中涉及的领域中的概念和分类, 如关联方交易的顶层概念为分类标准中的“关联方交易”(CAS36关联方披露[836600]附注—关联方关系及其交易)。

### 2.4.2 构建本体类和层次

#### (1) 本体类.

本体类是领域内一系列对现实世界进行抽象的相关元素集合. 根据财务报告关联交易披露的特点, 本文针对顶层概念“关联方交易”构建了关联交易领域本体类, 包括5个基本类、若干相应的二、三级子类. 实际应用中可根据领域概念特点对类与子类进行扩展(见表1)。

表1 关联方交易领域本体类(CAS36)

Tab.1 Domain ontology class of transaction in relevant parties

类别		类及其子类			
一级类	业务	要素	关系	业务区域	时间
二级类	财产保险、人身保险	保险代理、金额范围	直接或间接地控制其他企业或受其他企业控制,以及同受某一企业控制的两个或多个企业:合营企业、联营企业、主要投资者个人、关键管理人员或其关系密切的家庭成员、受主要投资者个人、关键管理人员或其关系密切的家庭成员直接控制的其他企业	区域A、区域B、区域C……	交易时间点交易 涵盖时间段
三级类	财产损失保险、责任保险、信用保险,人寿保险、健康保险、意外伤害保险等	保费收入、分出保费收入	**集团公司、**有限公司、**公司、** (人名)等	区域A <sub>1</sub> 、区域A <sub>2</sub> 、...、区域N <sub>n</sub> 等	**年度、**季度、**月**日等

首先,构建“业务”基本类,用于描述关联交易的分类,以保险业务为例,其下可以分为人寿保险、财产保险2个子类.按照功能类别每个子类还可再分为不同的三级子类,本文中的财产保险类可以再细分为对财产损失保险、信用保险和责任保险等.

其次,构建“要素”基本类,用于描述业务主要内容,其下分为保险代理、金额范围2个子类,再按照实际情况分别具体划分到保费收入、分出保费收入等.

再次,构建“关系”基本类,用于描述关联方的性质,其下可分为直接或间接地控制其他企业或受其他企业控制,以及同受某一企业控制的两个或多个企业、合营或联营企业、主要投资者个人、关键管理人员或其关系密切的家庭成员、受主要投资者个人、关键管理人员或其关系密切的家庭成员直接控制的其他企业等子类,再按照具体关联方名称向下分为企业或个人单位,例如\*\*集团公司、\*\*有限公司、\*\*公司、\*\* (人名)等.

然后,构建“业务区域”基本类以及相应的二级、三级类,用于描述业务开展的区域及子区域位置.

最后,构建“时间”基本类,用于描述业务发生时间以及业务涵盖的时间,再向下细分为\*\*年度、\*\*季度、\*\*月\*\*日等.

(2) 本体层次.

财务报告本体类的层级展示可借鉴分类标准的展示链接库的层级关系,将财务报告元素逐级展开表现出财务报告类的层次效果<sup>[16]</sup>.如图2所示.

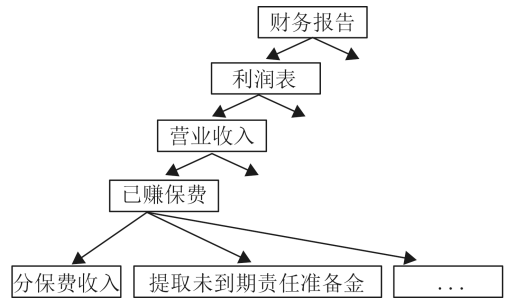


图2 财务信息本体类层级

Fig.2 Levels of financial information ontology class

2.4.3 类属性的定义

本体中的对象属性和数据类型属性可以对本体类的关系和限制进行描述,对象属性表示对象间的关系,数据类型属性则表示数据与对象类型值间的关系.财务会计领域中,定义基本类的对象属性为从属、合并或加总等元素.以保险行业基本类“业务”为例,参照XBRL分类标准,将“保险业务”的数据类型属性定义成:业务编码、业务名称等(详见表2业务属性栏目).一级类属性的定义可通用,子类属性的定义则需根据具体情况做出调整.以非寿险业务收入为例,保险期间与会计期间的存在不一致的现象,“根据权责发生制原则,对于保险风险未经过期间的收入应通过负债进行相应的调整;保险风险已经过期间的收入,即已赚保费才是真正意义上的保费收入,可导致保险人收到保险费的时间与保险风险时间不匹配”<sup>[20]</sup>.参照新的会计准则对于非寿险业务的描述,设置其属性为保险期间、责任期已过、责任期

表2 保险业务属性主题词表

Tab.2 Subject list of insurance business attributes

业务属性	属性类型	同义属性	近义属性	属性定义
业务编码	通用	业务编号	业务序号	编码是由数字、字母和连接符等符号按一定规则组成的字符串,用以唯一标识具体的业务
业务名称	通用	业务名字	名称	业务名称是指业务的名称,用于标识某一具体业务
业务类型	通用	业务种类	业务分类	业务类型是按照业务内容,按照企业会计准则描述的业务所属的业务类型
对应科目	保费收入	保险业务收入	再保险收入	企业会计准则对保费收入科目进行的定义
相关概念	保险合同	原保险合同 (原会计准则规定)	再保险合同 (新会计准则规定)	再保险合同,是指一个保险人(再保险分出人)分出一定的保费给另一个保险人(再保险接受人),再保险接受人对再保险分出人由原保险合同所引起的赔付成本及其他相关费用进行补偿的保险合同.

未过、保险风险等,其时间计量单位定义为年度、保险协议单位定义为人民币元,并设置相应的定义域和值域范围.

对信息资源中的属性和约束进行定义是语义集成的核心,根据属性和约束构建的主题词表为本体映射的提供依据.为了保证尽可能多的保险业务属性在主题词表中出现,可以将保险业务主题词表划分成业务属性、属性类型等5项(见表2).

### 2.4.4 一般领域本体构建

一般领域本体描述了领域内广泛使用领域中的概念、概念的属性、概念间的关系、概念的实例以及属性和关系的约束规则,详细描述各关联方交易特点及关键数据<sup>[21]</sup>.随着企业信息管理系统的发展,当前的企业财务数据往往以关系数据库的形式存储在财务数据库中,关系数据表中有着丰富的语义信息和主题词表,这些可以为自动快速构建领域本体提供可行路径.本文实验过程中就是把现有财务数据库中的财务报表科目作为一般领域本体构建的数据资源,通过模式映射将报表科目生成本体,再通过本体映射规范化,进一步得到领域本体.

### 2.4.5 概念语义关系构建

概念间关系是描述概念间的语义关系,概念间语义关系的M对N(M、N分别可以大于等于1)种关系形式化表达了不同层级类别的概念间的联系或约束,是实现查询和推理的基础.本体中最常用的关系

包括表达概念间继承关系和构成父类子类的层次结构的kind-of,以及表达整体一部分关系的part-of.除此之外,为表达更加丰富的语义,可以根据需要以Object Property(对象属性)的形式自定义概念间的语义关系.以保险企业关联交易知识本体为例,以关联交易文本内容结构为基础(见图3),定义关联交易知识本体概念关系集(见表3),以此定义概念间的语义关系(见图4).

表3 关联交易知识本体概念关系集

Tab.3 Ontology concept relationship collection of relevant transaction knowledge

属性名称	连接对象	具体含义
type	Business contract	描述关联交易类型
belongs to	notes to financial statements, belongs to	描述该关联交易所属财务报表附注内容
belongs to	Main business income, belongs to	描述该关联交易为主营业务收入子类
kind-of	Part_business	描述保险业务代理协议为关联交易所属子类
has business doc	Task, business_doc	描述交易的商务文档
has business with	opposite party, contract	描述本次交易对象
Date	date, contract	描述合同签订日期
duration	duration, contract	描述合同涵盖日期
sum of	sum of money, contract	描述合同金额
corresponding account	corresponding account, contract	描述合同对应科目为营业收入科目
corresponding account	corresponding account, contract	描述合同对应科目为应收账款/银行存款/其他

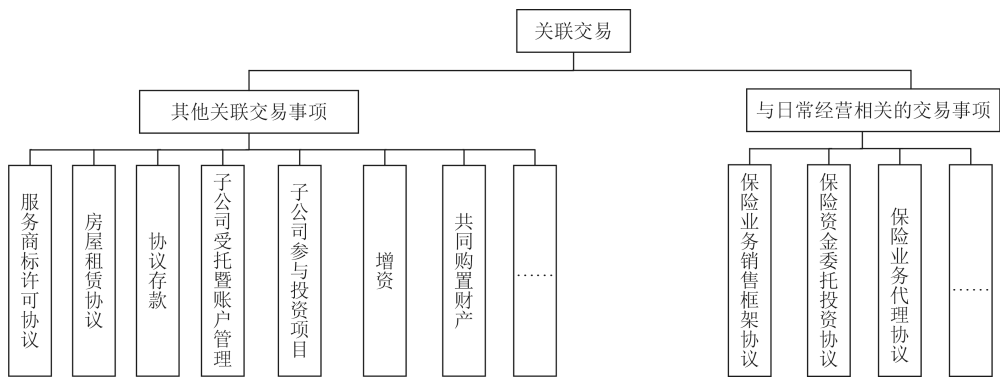


图3 关联交易文本内容结构图

Fig.3 Content structure of relevant transaction text

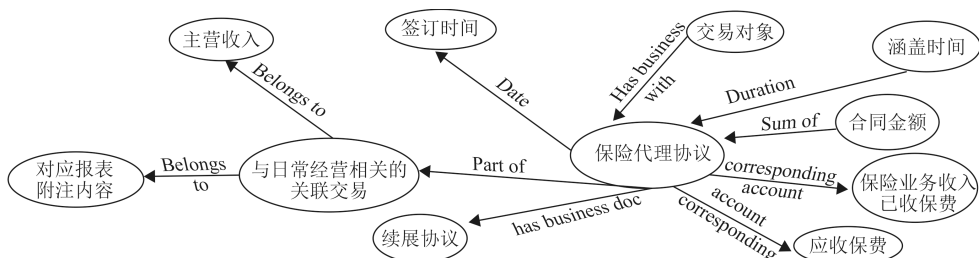


图4 关联交易内部概念间语义关系图

Fig.4 Semantic relationships among the relevant transaction inner concepts

针对财务报告中该段文本内容“本公司与集团公司于2011年12月15日签订2011年续展确认书,将上述保险业务代理协议进一步续展至2014年12月31日.在截至2014年12月31日的3个年度,该交易金额年度上限均为人民币11.88亿元.”以财务报告中文本表述与财务报告附注中表格表述的财务信息关

联性为切入点,将XBRL分类标准及元素作为财务术语表,以此为索引对文本进行自然语言处理(分词、命名实体识别等),根据短语相关性、语义相关性对文本中的财务信息进行标注并抽取,形成的结构化实例结果如图5所示.可参照上文就续展协议进行内部层级语义关系进一步展开构建,此处不再赘述.

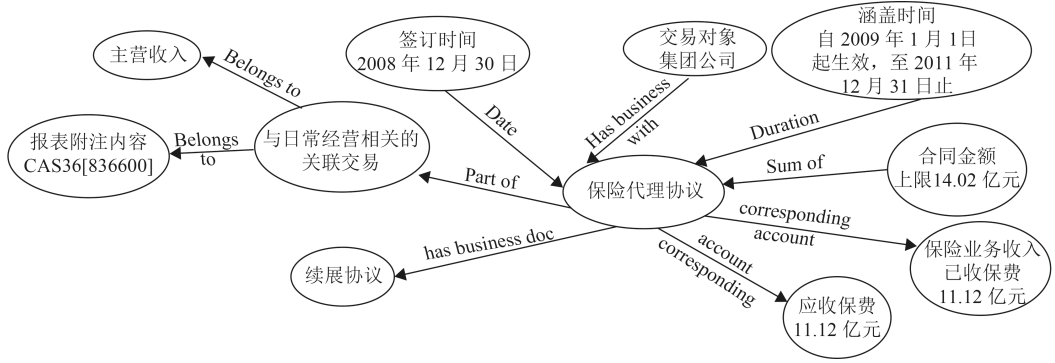


图5 概念间语义关系实例图

Fig.5 Instance of semantic relationships among concepts

### 3 文本结构化处理与XBRL转换

根据上述构建流程,将上例文本内容结构化处理之后使用富士通InterstageXWand Taxonomy Editor / Instance Creator将其转换成XBRL数据.

(1) 创建扩展分类标准. 首先定义XBRL通用分类标准中没有定义的需要扩展的财务元素及其元素属性,并在此基础上定义分类标准的定义、标签、展

示、计算及参考链接库,以本例中标签、展示两个链接库为例,可以得到扩展分类标准.

(2) 创建XBRL实例文档. 在扩展分类标准基础上,将财务报告事实值与扩展分类标准中对应元素进行映射,同时确定上下文、单位,生成保险业务代理协议XBRL实例文档. 图6、7分别为上述文本内容进行信息抽取和处理后形成的实例文档、实例文档代码以及HTML形式的报表.

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Generated by Fujitsu Interstage XWand Taxonomy Editor -->
<xbrli:xbrl xmlns:ins="http://www.fujitsu.com/xbrl/taxeditor/insuranceagencyagreement"
xmlns:link="http://www.xbrl.org/2003/linkbase"
xmlns:iso4217="http://www.xbrl.org/2003/iso4217"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xbrl="http://www.xbrl.org/2003/instance"
xmlns:xlink="http://www.w3.org/1999/xlink">
<link:schemaRef xlink:type="simple" xlink:href="insuranceagencyagreement.xsd"/>
<xbrli:context id="cas36_20101231">
<xbrli:entity>
<xbrli:identifier scheme="http://www.sse.com.cn/">601628</xbrli:identifier>
</xbrli:entity>
<xbrli:period>
<xbrli:instant>2010-12-31</xbrli:instant>
</xbrli:period>
</xbrli:context>
<xbrli:unit id="CNY">
<xbrli:measure>iso4217:CNY</xbrli:measure>
</xbrli:unit>
<ins:TradeObject contextRef="cas36_20101231">集团公司</ins:TradeObject>
<ins:CoverageTime decimals="0" contextRef="cas36_20101231"
unitRef="CNY">3</ins:CoverageTime>
<ins:ContractAmount decimals="0" contextRef="cas36_20101231"
unitRef="CNY">1402000000</ins:ContractAmount>
<ins:SigningTime contextRef="cas36_20101231">2008-12-30</ins:SigningTime>
<ins:ReceivablePremium decimals="0" contextRef="cas36_20101231"
unitRef="CNY">1112000000</ins:ReceivablePremium>
</xbrli:xbrl>
```

图6 保险业务代理协议实例文档代码片段

Fig.6 Code of insurance agency protocol business instance

### 4 总结与展望

本文探讨了上市公司年报中非结构化信息的获取方法,提出了使用知识本体方法将非结构化的文

本信息向结构化文本信息进行转换,并最终形成XBRL格式以提高披露信息中文本信息处理效率的方法.为了验证此方法的可行性,本文以某保险公司

All Contexts	
保险业务代理协议	
交易对象	集团公司
涵盖时间	3
签订时间	2008-12-30
合同金额	14 020 000 000
应收保费	11 120 000 000
已收保费	
与日常经营相关的关联交易	

图7 关联交易文本信息抽取后形成的HTML报表

Fig.7 Reporting in HTML after information extracting to relevant transaction text information

2010年年报中关联交易的保险业务代理协议文本块为例,构建了财务报告领域本体,探讨了财务报告文本信息抽取方法,最后利用XBRL表示了结构化的信息,该方法对实现利用计算机技术高效、准确分析财务报告中的非结构化信息有参考意义。

#### 参考文献:

- [1] LI H Q, ZHAI J. Literature review of XBRL semantic research[C]// 2015 International Conference on Computer Science and Intelligent Communication. HK: Atlantis, 2015: 316-320.
- [2] LI M J, ZHOU Z H, DU M J. Detection and resolution of structural conflictions in heterogeneous XBRL taxonomies[C]// The 5th International Conference on New Trends in Information Science and Service Science. HI: IEEE, 2011: 312-317.
- [3] LI M J, ZHOU Z H, DU M J. XBRL in the Chinese financial ecosystem [J]. IT professional, 2013, 15(6): 36-42.
- [4] 李吉梅, 杜美杰. 基于XBRL的异构财务信息集成算法[J]. 吉林大学学报(工学版), 2012, 42(S1): 266-270.
- LI J M, DU M J. Information integration algorithm of heterogeneous XBRL financial reporting [J]. Journal of Jilin University(Engineering and Technology Edition), 2012, 42(S1): 266-270.
- [5] PAN D, PAN Y S. Incorporating XBRL into business intelligence applications based on formal semantics[C] // 2011 China Academic Accounting Association Annual Meeting. XM: Elsevier, 2011: 1758-1765.
- [6] 冯志伟. 现代术语学引论(增订本)[M]. 北京: 商务印书馆, 2011. 12-195.
- [7] 杨周南, 朱建国, 刘锋. XBRL分类标准认证的理论基础和方法学体系研究[J]. 会计研究, 2010, 1(11): 10-15.
- YANG Z N, ZHU J G, LIU F. Research on the theory basis and methodology system of xbrl taxonomy recognition [J]. Accounting Research, 2010, 1(11): 10-15.
- [8] DEBRECENY R, FELDEN C, OCHOCKI B, *et al.* XBRL for interactive data[M]. NY: Springer, 2009. 189-211.
- [9] LARA R, CANTADOR I, CASTELLS P. XBRL taxonomies and OWL ontologies for investment funds[C]// In the 1st International Workshop on Ontologizing Industrial Stand-

- ards at the 25th International Conference on Conceptual Modelling. AZ: Springer, 2006: 271-280.
- [10] BAO J, RONG G, LI X, *et al.* Representing financial reports on the semantic web: a faithful translation from XBRL to OWL[C]//International Workshop on Rules and Rule Markup Languages for the Semantic Web. DC: Springer, 2010: 144-152.
- [11] HUANG M, WANG D, WANG K. Ontology-based semantic retrieval of XBRL data[C]// 2011 International Conference on Business Computing and Global Informatization, SH: IEEE, 2011: 363-366.
- [12] ZHU H. Semantic integration approach to efficient business data supply chain: integration approach to interoperable XBRL[EB/OL]. (2007-10-01)[2016-04-01]. <http://web.mit.edu/smadnick/www/wp/2007-10.pdf>
- [13] ROMILLA C, YOON VY, REDMOND RT, *et al.* Ontology based integration of XBRL filings for financial decision making [J]. Decision Support Systems, 2014, 1(68): 64-76.
- [14] GARCIA R, GIL R. Publishing XBRL as linked open data[C]// In Proceedings of World Wide Web Workshop: Linked Data on the Web, Madrid: CEUR-WS, 2009: 538
- [15] KAMPGEN B, WELLER T, O'RIAIN S. Accepting the XBRL challenge with linked data for financial data integration [J]. Lecture Notes in Computer Science, 2014, 1(8465): 595-610.
- [16] 吴忠生, 张天西, 陈志德. 基于领域本体的XBRL财务报告转换研究[J]. 计算机应用研究, 2013, 1(30): 3643-3646.
- WU Z S, ZHANG T X, CHEN Z D. Research on conversion between XBRL financial reports based on domain ontology [J]. Application Research of Computers, 2013, 1(30): 3643-3646.
- [17] ANTONINA K, CAMILLA M, BARBRO B. Mining textual contents of financial reports [J]. The International Journal of Digital Accounting Research, 2004, 4(7): 1-29.
- [18] MENDEZ NUNEZ S, TRIVIO G. Combining semantic web technologies and computational theory of perceptions for text generation in financial analysis[C]// 2010 IEEE International Conference on Fuzzy Systems. Barcelona: IEEE, 2012: 1-8.
- [19] GRUBER T R. Toward principles for the design of ontologies used for knowledge sharing [J]. International journal of human-computer studies, 1995, 1(43): 907-928.
- [20] 李群. 非寿险业务的会计核算[J]. 财务与会计, 2009, 1(5): 20-26.
- LI Q. Accounting for non-life insurance business [J]. Financial and Accounting, 2009, 1(5): 20-26.
- [21] 黄蓉, 徐璐璐. 公司关联交易文献评述[J]. 广东工业大学学报, 2016, 33(06): 102-106.
- HUANG R, XU L L. Summarization of related party transactions in listed company [J]. JOURNAL OF GUANGDONG UNIVERSITY OF TECHNOLOGY, 2016, 33(06): 102-106.