

# 基于混合本体的保险公司数据集成研究

孙建光, 刘 英

(大连海事大学 交通运输管理学院, 辽宁 大连 116023)

**摘要:** 参照“Mediator/Wrapper”体系结构, 利用混合本体的概念设计了基于本体的保险企业数据集成模型, 以六元组的方式为例完成了对保险公司新型人寿保险信息的混合本体定义。利用本体技术通过领域建模描述该领域的概念及其相互关系, 可以有效地完成保险公司异构数据的集成和深层的数据共享, 提供统一的信息查询及用户视图, 进而提高保险公司管理支持和决策的可靠性和准确性。

**关键词:** 异构数据; 语义异构; 数据集成; 本体; 人寿保险

中图分类号: TP391

文献标识码: A

文章编号: 1674-7720(2010)16-0087-03

## Insurance data integration research based on hybrid ontology

SUN Jian Guang, LIU Ying

(School of Transportation Management, Dalian Maritime University, Dalian 116023, China)

**Abstract:** With the reference of “Mediator/Wrapper” architecture together with hybrid ontology theory, an ontology based insurance data integration model is proposed, with an example of the hybrid ontology definition given under the six-point group. The model can benefit the insurance company about the heterogeneous data integration for their data sharing, integrated inquiry and full view of customer, and it will finally improve the capability of the management decision.

**Key words:** data heterogeneity; semantic heterogeneity; data integration; ontology; life insurance

过去二十多年来中国保险企业的信息化建设大多缺乏统筹规划, 造成了各种不同应用系统盲目上马的现象。不同的业务系统使用不同的后台数据库, 不同的系统开发人员对业务术语缺乏统一定义, 这些问题引发了现今企业系统管理的巨大问题; 同时由于大多数应用系统之间缺乏标准化的数据接口定义, 因此不同的应用系统之间必然会成为彼此隔离的信息孤岛, 信息孤岛是当前信息化建设中亟需解决的主要问题<sup>[1]</sup>, 彻底消除信息孤岛, 有效地集成现有及未来的业务应用系统的信息数据已成为当前信息化建设的重点工作。

### 1 保险公司数据集成研究

数据量大和数据安全要求高是保险业的特点, 中国保险业发展到今天, 大多数人寿保险公司除具备核心业务系统之外, 还建立了办公自动化系统、财务系统、决策支持系统、数据仓库以及相关的其他系统, 各种类型的数据在成倍增长。目前来看, 主流保险公司的数据集中已告一段落, 接下来的主要课题是数据的共享及集成利用。

国外许多保险公司早在十几年前就开始了数据集成

技术的研究和运用, 多数保险公司已完成了数据集成和整合技术对其个险系统、团险系统、银行保险系统、年金系统和养老金系统等用户及保单数据的集成<sup>[1]</sup>。

在业界, 数据集成的概念比较混乱, 例如系统整合、应用整合、展现整合、存储整合、数据库整合、数据大集中等。这些不同的概念从不同的层次、不同的角度阐述了信息系统整合的内涵和外延。本文主要探讨异构平台下的数据集成。

一般情况下数据的异构包括两种, 即语法异构和语义异构。语法异构包含不同的数据库系统或不同数据结构, 对于此类异构数据主要的集成方法是将异构数据转换成XML数据模式。对于语义异构的数据集成可以利用本体技术通过领域建模描述该领域的概念及其相互关系。

### 2 本体的概念及建模

#### 2.1 本体论的概念

起源于哲学的本体论源于对万物本质的追问, 其派生于希腊语 onto 和 logia, 是一门研究事物根本的形而上的科学。本体论在哲学外的应用主要集中在信息技术和

人工智能领域。伴随着领域建模、知识工程和数据库等技术的发展,在各领域中需要一个通用的概念描述,来说明在该领域本质上存在的对象、过程、属性及相互依存关系等。

1980年 McCathy 基于前人的理论提出“以逻辑概念为基础的智能系统必须列出所有存在的事物并构建一个本体来描述我们的世界”,正式提出人工智能领域的本体论概念<sup>[2]</sup>。1993年 Gruber 提出第一个被IT领域广泛接受的本体论正式定义“an ontology is an explicit specification of a conceptualization”<sup>[3]</sup>。1998年 Guarino 通过分析本体论与概念化之间的区别对 Gruber 的定义进行修订并提出“域空间”的概念,在此基础上通过在澄清本体、预定模型和概念化的基础上得到了本体论的定义:“本体论是一个逻辑理论,用来说明一个正规词汇表的预定含义。”<sup>[4]</sup>

## 2.2 本体建模

在设计本体之前,首先要进行本体建模,也就是用形式化方法定义本体的语义,本体的建模是知识工程中实现知识重用和语义共享的基础,到现在还没有一种统一的形式化本体定义标准。

研究者们根据研究与应用的不同背景给出了多种形式化的定义方式,六元组定义<sup>[5]</sup>是比较流行的一种,其形式化定义如下:

定义1 资源库本体是一个六元组:

$$O = \{C, A, R, H, I, X\}$$

其中,  $C$  是概念的集合;  $A$  是属性的集合;  $R$  是关系的集合;  $H$  是概念层次;  $I$  是实例的集合;  $X$  是本体公理的集合。

## 3 基于本体的保险公司数据集成实施模型

### 3.1 基于本体的保险公司数据集成模型

在大多数的大型企业里,异构数据的集成非常复杂且对企业相当重要。来自经济和法规方面的压力使得这些企业更多地聚焦于如何获取和组织这些数据以进行集成,关注集成数据的质量和定义的标准。为此,企业开始注重通过数据集成管理来建好企业的数据集成模型,一个典型的基于本体的企业数据集成模型如图1所示。

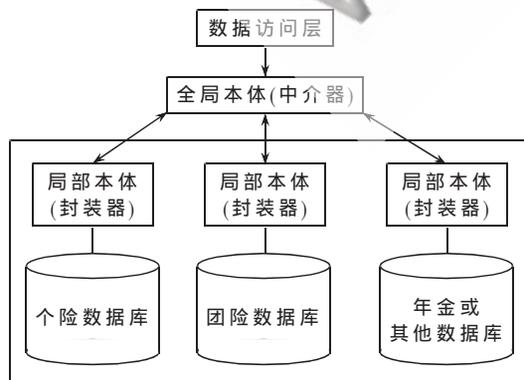


图1 基于本体的保险数据集成模型

上述基于混合本体的保险公司数据集成模型参照“Mediator/Wrapper”<sup>[6]</sup>体系结构,运用了混合本体的概念设计,将 mediator/wrapper 虚拟整合机制整合到系统中,通过 mediator 和 wrapper 来整合。mediator 和 wrapper 都是软件组件,位于用户和数据源之间,mediator 服务于处理用户提问和查询结果的整合,wrapper 则负责对信息源的连接和具体查询。

遵循 mediator/wrapper 整合机制可保持各个异构保险数据源的自治性,满足局部的各种应用,并同时发挥 mediator 的作用,满足全局性应用需求。在 mediator 中引入 ontology 等语义相关技术后,能够有效解决知识整合、个性化服务等问题。使用这种机制的整合系统不需要在本地存储大量资源,因而能够适应网络环境下信息源系统高度自治、数量多、更新频繁等特点。

### 3.2 混合本体的构建

举例说明基于混合本体的信息集成技术,对于不同的数据源(如个险系统和银行保险系统),将新型人寿保险信息按类型和销售渠道分类存放,构造局部本体  $O_1$  和  $O_n$  结构分别如图2、图3所示。

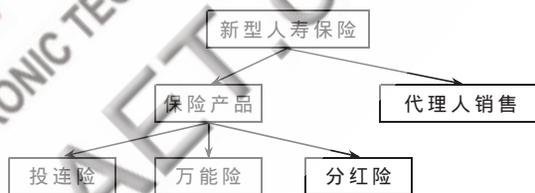


图2 局部本体  $O_1$

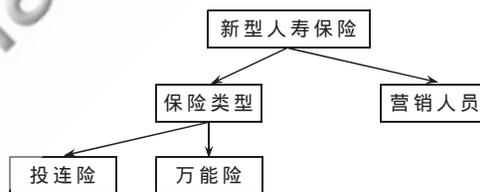


图3 局部本体  $O_n$

通过对数据源进行分析识别,找出两个不同的数据源中重要的原语来定义本体。

采用前面提到的六元组来完成  $O_1$  本体的建模,由于系统比较简单,只需要使用  $C$ 、 $A$ 、 $H$  三个元素,表示为:

$C = \{\text{新型人寿保险, 保险产品, 代理人销售人员, 投连险, 万能险, 分红险}\};$

$A = \{\text{包含(代理人销售, 保险产品), 归类于(保险产品, 新型人寿保险), 销售渠道(代理人销售人员, 新型人寿保险)}\};$

$H = \{\text{(投连险, 保险产品), (万能险, 保险产品), (分红险, 保险产品), 保险产品}\dots\}$

用 OWL 语言描述本体  $O_1$ :

```
<owl: Class rdf:ID = "新型人寿保险">
```

```
<owl: Class rdf:ID = "保险产品">
```

```
<owl: Class rdf:ID = "代理人销售人员">
```

```

<owl: Class rdf:ID = "投连险">
  - <rdfs: subClassOf>
    <owl: Class rdf:about = "# 保险产品" />
  </rdfs: subClassOf>
</owl: Class >
<owl: Class rdf:ID = "万能险">
  - <rdfs:subClassOf >
    <owl: Class rdf:about = "# 保险产品" />
  </ rdfs:subClassOf>
</owl: Class >
<owl: Class rdf:ID = "分红险">
  - <rdfs:subClassOf>
    <owl: Class rdf:about = "# 保险产品" />
  </rdfs:subClassOf>
</owl: Class>
<owl: ObjectProperty rdf:ID = "包含">
  <rdfs: domain rdf:resource = "# 代理人销售人员" />
  <rdfs: range rdf:resource = "# 保险产品" />
</owl: ObjectProperty>
<owl: ObjectProperty rdf:ID = "归类于">
  <rdfs: domain rdf: resource = "# 保险产品" />
  <rdfs: range rdf: resource = "# 新型人寿保险" />
</owl: ObjectProperty>
<owl: ObjectProperty rdf:ID = "销售渠道">
  <rdfs:domain rdf:resource = "# 代理人销售人员" />
  <rdfs:range rdf:resource = "# 新型人寿保险" />
</owl: ObjectProperty>

```

参照以上定义可以得到本体  $O_{II}$  的描述。分析可得局部本体  $O_I$ 、 $O_{II}$  存在以下问题:

- (1) 使用不同的元语表示同一概念:  
 保险产品 → 保险产品;  
 代理人销售人员 → 营销人员;  
 投连险 → 投资联结保险

- (2)  $O_I$  中包含了  $O_{II}$  中不存在的概念“分红险”。

通过对类、属性及其类间关系的集成合并局部本体, 将  $O_I$ 、 $O_{II}$  两个本体联系起来建立一个全局本体, 然后在全局本体和局部本体之间建立映射, 可得到如图 4 所示的全局本体  $O_G$ 。

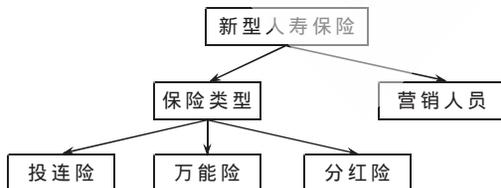


图 4 全局本体  $O_G$

共享词汇集为: 新型人寿保险, 保险产品, 营销人员, 投连险, 万能险, 分红险。全局本体  $O_G$  中的“保险产品”是由  $O_I$  中的“保险产品”和  $O_{II}$  中的“保险产品”间相等的概念合并得到。全局本体  $O_G$  中的概念“分红险”

是由  $O_I$  中的“分红险”直接拷贝得到的。其相应的三元组为:

$C = \{ \text{新型人寿保险, 保险产品, 营销人员, 投连险, 万能险, 分红险} \};$

$A = \{ \text{包含(营销人员, 保险产品), 归类于(保险产品, 新型人寿保险), 销售渠道(营销人员, 新型人寿保险)} \};$

$H = \{ (\text{投连险, 保险产品}), (\text{万能险, 保险产品}), (\text{分红险, 保险产品}), \text{保险产品} \dots \}。$

全局本体在混合本体中的功能为:

(1) 在异构数据源和应用界面之间形成一个中介层, 由于数据来源的复杂性, 这些数据可能存放在不同的地理位置、不同的数据库和不同的应用之中;

(2) 利用共享概念创建全局本体;

(3) 提供给用户界面的综合查询一个概念以上的统一视图, 用户通过提交一个基于全局本体之上的 RDF 查询就能获取所有相关数据源的数据以实现概念上的互操作。

通过参照“Mediator/Wrapper”体系结构, 利用混合本体的概念设计基于本体的保险企业数据集成模型, 可以有效地完成保险公司数据大集中后的异构数据的集成和深层的数据共享, 提供统一的信息查询及用户视图, 进而提高保险公司管理支持和决策的可靠性和准确性。可以看到通过成体系的语义分析完成保险公司通用的本体构建并对其进行评价, 以及对映射规则及其推理引擎的完善等问题还需要进一步研究。

#### 参考文献

- [1] RADCLIFF J. Integrate your data to create a single customer view[OL]. Gartner, 2004.
- [2] CARTHY J M. Circumscription—a form of non-monotonic reasoning[J]. Artificial Intelligence, 1980, 5(13): 27-39.
- [3] GRUBER T R. Towards principles for the design of ontologies used for knowledge sharing. Stanford University, Tech Rep: KSL-93-04, 1993.
- [4] 李善平, 尹奇韡, 胡玉杰, 等. 本体论研究综述[J]. 计算机研究与发展, 2004, 41(7): 1041-1052.
- [5] HARRY R L, CHRISTOS H P. Elements of the theory of computation(second edition)[M]. Prentice Hall PTR Upper Saddle River, NJ, USA, 1997.
- [6] 卓国锋, 罗军. 基于 Mediator/Wrapper 信息集成的查询优化研究[J]. 计算机工程与应用, 2007, 43(12): 159-161, 242.

(收稿日期: 2010-04-21)

#### 作者简介:

孙建光, 男, 1978年生, 博士研究生, 主要研究方向: 金融行业企业信息技术管理。