

面向语义信息共享的元数据模型的研究与实现^{*}

黄宏斌 张维明 邓 苏 刘 震

(国防科技大学信息系统与管理学院 长沙 410073)

摘 要 元数据在国内依然是一个研究热点,本文面向基于语义的信息资源共享就如何利用人工智能领域内的本体(ontology)理论,建立用于描述信息资源语义及其之间关系的元数据模型进行了研究,并初步实现了基于元数据模型的信息资源元数据描述模版和描述工具原型。

关键词 元数据模型,本体,语义,元数据规范模版,语义覆盖网

Research and Implement of MetaData Model for Information Semantics Sharing

HUGAN Hong-Bin ZHANG Wei-Ming DENG Su LIU Zhen

(College of Information System and Management, National University of Defense and Technology, Changsha 410073)

Abstract Metadata is still a hot topic for information sharing. It usually does not support information semantics management. This paper, orienting to semantics-based information resource sharing, focuses on the design of the integrative metadata model by using ontology theory in AI domain. The metadata model is used to describe the semantic of information resource. And use the metadata to implement the metadata specification template of information resource and the prototype of metadata description tools.

Keywords Metadata model, Ontology, Semantic, Metadata specification template, Semantic overlay network

1 引言

随着网络的发展和信息时代的到来,在一些大型企事业单位和虚拟组织环境中,存在着大量的业务信息系统和信息资源,组织内的各单位依据业务或专业知识或任务进行分布,并且管理着一组信息资源集合。不同单位有着不同的工作重心和目标,因此会在不同工作领域下对底层的信息资源采用不同的概念实施建模。这些分布在广域网络环境下的信息资源具有动态、分布、多元和无序等特点,而信息共享与服务体系尚未建立,缺乏统一、科学、高效的元数据体系。各部门或组织之间的信息相对封闭,形成了一个“信息孤岛”,造成了巨大信息资源浪费,同时信息资源的异构性为信息之间的共享与互操作带来了困难。对用户而言,访问并有效利用这些信息资源,实现各节点分布的信息资源的共享和交换,以及分布信息系统之间的互操作,都迫切需要一套完整的、基于语义的准确规范和可操作的元数据模型及其描述规范来支持。

元数据是关于数据的数据,是信息资源组织管理与共享服务的基础之基础,重中之重。元数据在国内依然是一个热门概念,是信息资源组织、存储、管理、发现、检索与利用的必要基础。元数据模型是针对元数据的统一建模,为分布、异构信息资源的共享提供了一个有效工具。基于本体的元数据通过本体来表示元素模型,将有效地促进领域知识的共享和语义表达,能够为用户和应用提供语义查询和信息汇集能力。但是,现有基于本体的信息共享与服务系统中几乎都假设共享一个集中的全局本体,用来访问信息资源,这样的维护开销比较大。在广域网络环境下,这种假设是不存在的。在网络环境中,每一个结点仅仅维护着本地信息资源的局部领域视

图,通过网络中结点的协作来发现、理解和使用领域范围内分布、异构、不断变化的信息资源。

本文从现实需求出发,针对在广域分布环境下进行信息共享与服务的需要,针对基于语义覆盖网的信息服务要求设计了面向语义信息共享的元数据模型,给出了一种面向语义的可扩展元数据框架;结合面向信息资源语义内容的描述和信息资源访问的管理元素描述,对元数据规范进行了设计,并初步实现了基于语义的元数据可视化描述工具。

本文的组织结构如下:第1节主要介绍了元数据模型的提出的背景,第2节面向用户需求的信息服务体系框架以及元数据模型需求描述。第3节主要介绍元数据模型框架及元数据模型定义。第4节基于元数据模型提出元数据模型描述规范模版。最后以一个元数据描述工具运行示例结束本文。

2 问题需求描述及思路

本文提出的元数据模型是面向一种以信息资源中心(information resource center, IRC)为基础的语义覆盖网,基于语义覆盖网^[3]对分布异构信息资源进行管理。各信息资源中心在技术上是相互独立的,系统各信息资源中心互联形成信息资源中心(IRC)网络,因此信息资源中心(IRC)网络类似一个广域分布环境下的对等网。语义覆盖网的目标是连接分布环境中各个信息孤岛,实现信息科学高效和安全地共享,元数据是发现、组织和使用分布环境中各类信息资源的基础。

2.1 问题需求描述

本文侧重于元数据模型的研究。针对分布异构环境的信息资源的特点,对元数据框架的设计实现必须满足如下需求:

1) 统一的标准化元模型 环境中的信息资源具有格式多

^{*}国家自然科学基金资助项目(批准号:60172012)。黄宏斌 讲师,博士研究生,主要研究方向为信息管理与智能决策支持技术;张维明 教授,博士生导师,主要研究领域为信息管理与辅助决策;邓 苏 教授,主要研究领域为信息管理、智能决策支持技术。

样、异构性强的特点,提供统一的、标准化的元数据模型是实现信息互联互通互操作的基础;

2)元数据模型要兼顾通用性和特殊性 环境中信息资源具有涉及领域多、分布地域广的特点,元数据标准的设计必须考虑其在一定范围内的通用性,使其尽可能地覆盖多种相似或相近的对象实体,同时兼顾能适应不同领域、不同特色和不同团体的特殊性要求;

3)全局可用性 用户或应用可以通过元数据及元数据知识库发现信息资源的存在,并且可以在授权范围内通过虚拟链接或本地缓冲对具体关注的信息资源进行访问;

4)面向语义 环境中信息资源复杂多样,为了保证信息资源的可理解性,应提供面向语义的信息资源描述能力,使各类用户或具体应用能够从结构和语义上理解信息资源,易于判断和决定数据如何使用,来满足特定需求;

5)可扩展性 环境中管理的信息资源具有动态性特征,因而要保证元数据标准设计的可扩展性;

6)支持互操作性 提供一种多对多的、松耦合的信息交互机制,支持的交互接口可以是预定义的,也可以是预先未知的。同时,在各局部元数据模式之间要能形成一定映射和进行语义相似度度量。

2.2 元数据框架

基于以上元数据需求,为了保证共享环境中资源的全局可视性、可理解性、可用性和互操作性,信息资源对象元数据遵循特定的元数据规范模板(Metadata Specification Template, MST)进行描述,元数据模板规范是基于元数据模型对信息资源对象进行描述的规范和要素。

如图1所示,整个元数据框架分为三个层次:第一层是元数据模型层,它是规范设计定制某类信息资源所用的 MST 时需要遵循的规则和方法,它是抽象化的元数据,从更高层次上规定了元数据的功能、数据结构、格式、设计方法、语义语法规则等多方面内容。第二层是元数据规范模板(MST)层,它是描述某类资源的具体对象时所有规则的集合,定义了元数据模式。不同领域、不同类型的资源可能会有不同的 MST,它包括了完整描述一个具体信息资源对象时所需要的数据项集合、各数据项语义定义、规则和计算机应用时的语法规则。各类专业或非专业编目人员遵循特定的 MST 对具体信息资源对象的元数据进行描述。第三层是元数据记录层,元数据记录是基于数据模式建立起来的应用元数据,是描述一个具体信息资源对象,并能对这个对象进行定位、管理,且有助于它的发现、获取和使用的数据。最终用户通过资源元数据来发现和使用信息资源对象。

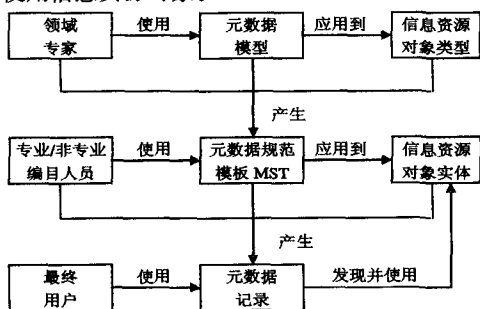


图1 元数据框架的层次关系图

3 元数据模型

元数据作为描述某种类型资源属性、提供信息资源定位、

管理和精确检索服务的结构化数据,是一种可以被机器理解和处理的信息。规范、标准的信息资源元数据结构是构建信息资源共享与服务平台的基础^[6]。元数据模型是规范设计定制某类信息资源所用的 MST 时需要遵循的规则和方法,它是抽象化的元数据,从更高层次上规定了元数据的功能、数据结构、格式、设计方法、语义语法规则等多方面内容。采用本体的设计思想来描述元数据,可以为元数据添加语义信息,从而能实现基于语义的元数据集成与共享。本文的元数据模型是基于本体的思想描述的。本文中的每个 MST 都采用一种基于本体的思想进行形式化描述,相应 MST 是建立在相关领域理解的基础上。因此在介绍元数据模型之前首先介绍本体的相关概念。

3.1 本体概述

本体在元数据描述和使用的语义化中起着非常重要的作用,主要体现在:1)本体不但能够保证以往基本的元数据描述能力,同时是语义信息描述的基础。语义信息主要由语义类、语义属性、语义关系、语义规则和语义实例构成。这正好对应着本体中的本体概念、概念属性、概念关系、本体规则和公理以及本体实例。2)本体将有效扩展信息检索能力。由于本体描述对各类信息资源赋予了语义,同时本体自身具有一定的推理能力,从而可以利用本体提供面向语义的查询能力,同时可以利用本体进行查询扩充,从而使检索的结果更加全面。3)本体为信息资源提供了一种面向语义的输出能力和相似度计算能力,包括元数据定义的语义输出和信息资源语义实例的语义输出,从而使用户能够更好地理解所需的信息资源。

本体(Ontology)最早是哲学领域的一个概念。从哲学范畴来说,本体是客观存在的一个系统的解释或说明,关心的是客观事物存在的抽象本质。牛津英语辞典里对“ontology”的解释是“对于存在的研究或科学(the science or study of being)”。自上世纪90年代以来,Ontology被广泛应用于人工智能、计算机等领域。关于Ontology的定义有许多,目前获得较多认同的是1998年R. Studer的解释^[11]：“Ontology是对概念体系的明确的、形式化的、可共享的规范说明”。由此定义,我们可以看出,Ontology具有以下特征:①Ontology是一个概念模型,它是通过抽象出客观世界中的一些现象的相关概念而得到的模型,这个概念模型可以被理解和表示为概念的集合(如实体、属性等),以及它们的定义和相互间的关系。②Ontology所使用的概念的约束条件都有明确的定义。③Ontology所体现的知识是获得共同认可的知识,反映的是相关领域中公认的概念集。④Ontology作为一种能在语义层次上共享知识的概念模型,具有良好的概念层次结构和对逻辑的支持能力。

在知识工程领域,本体是由用于描述某种确定现实情况的特定术语集,加上一组关于术语内涵意义的显式的假定集合构成。在最简单的情况下,本体只描述概念的分类层次结构;在复杂的情况下,本体可以在概念分类层次的基础上,加入一组合适的关系、公理、规则来表示概念之间的其它关系,约束概念的内涵解释。这里我们可以将本体定义如下:

定义3.1 一个完整的本体应由概念、关系、函数、公理和实例等五类基本元素构成。因此,我们把本体表示为如下形式: $O = \{C, R, F, A, I\}$,其中

C:概念。本体中的概念是广义上的概念,它除了包括一般意义上的概念外,还包括任务、功能、行为、策略、推理过程等等。本体中的这些概念通常按照一定的关系形成一个层次

结构。

$R \subseteq 2^{C \times C}$: 概念之间的关系, 表示概念之间的一类关联。如概念之间的“subclass-of”关系、“part-of”关系等。

$F \subseteq R^n$ 是一种特殊的关系, 其中第 n 个元素 C_n 相对于前面 $n-1$ 个元素是惟一确定的。函数可以用如下形式表示: $F: C_1 \times C_2 \times \dots \times C_{n-1} \rightarrow C_n$ 。

A : 概念或者概念之间的关系所满足的公理, 是一些永真式。

I : 为领域内概念实例的集合。在有些本体模型中, 概念的实例不被看成是本体的组成部分。

3.2 元数据模型

本体(ontology)是领域知识的概念化说明, 它将特定领域有关的对象、概念及其关系以形式化的说明来严格规定。利用本体建立面向语义的元数据模型, 可以将元数据中实体类的含义、类间的关系更加明确地表达出来, 从而支持广域分布环境下的概念建模、信息搜索与交换、信息资源共享与服务系统建设等研究。而本文研究一种面向语义信息共享的信息资源元数据模型, 因此本文所提出的元数据模型是基于本体的一种元数据模型。

假设 T 为领域语义词典, 是领域术语词汇的集合, 对应领域语义词典中的词汇, 主要包括类术语、属性术语、关系术语等, D_{basic} 是预定义的基本数据类型集合, D_{enum} 是预定义的枚举类型。

领域语义词典用来规范 MST 中元素命名, 并定义了不同词汇间的语义关系。领域语义词典(Domain semantic dictionary, DSD)定义为一个受限的词汇知识库, 其面向的是领域的常用词汇和专用词汇, 其中定义了不同词汇间的语义关系。其目标是: 1) 规范不同信息领域中 MST 定义时信息资源描述的词汇, 从而可以有效地限制任意词汇描述引发的不一致性; 2) 是资源匹配过程中词汇语义相似度计算的基础。DSD 主要采用《知网》的语义描述结构来进行组织, 主要对实词进行描述, 是一种开放结构, 可以进行扩充。

我们基于信息资源都是信息单元组成的这一理解, 采用信息单元实体类和实体类之间的关系及相关约束和规范描述信息资源。具体定义如下: :

定义 3.2 元数据模型是一个八元组 $MD := \langle E, A, L, H^c, R, I, F, P \rangle$ 。其中

1) E 是信息单元实体类集合, $\forall c \in E, c = (Name, A^c)$, $Name(c)$ 表示 c 的命名词汇, $Name(c) \in T$, 其中 $A^c = \{x | x \in A, \text{且 } Att(x) = c\}$ 。 Att 是函数集中的属性映射函数。

2) A 是定义到信息单元实体类的属性集合, 属性又分为基础属性和复合属性。 $\forall a \in A, a = (Name, dt)$, $Name(a)$ 表示 a 的命名词汇, $Name(a) \in T, dt \in D_{basic} \cup D_{enum}$ 。

3) L 取值域集, $L = D_{basic} \cup D_{enum}$ 。

4) H^c 是实体类间的一种二元层次关系, 该关系是一种有向的、传递的关系, 是一种偏序关系, 包括了实体类之间的继承 is-a 关系和聚合(组合) part-of 关系。 $H^c \subseteq E \times E, H^c = H^{inh} \cup H^{Agr}$, $H^{inh}(c_1, c_2), c_1, c_2 \in E$, 表示 c_2 是 c_1 子类即 c_2 is-a c_1 , $H^{Agr}(c_1, c_2), c_1, c_2 \in E$, 表示 c_2 是 c_1 的组成部分即 c_2 is part-of c_1 , 如: H^{inh} (学员、本科生), H^{Agr} (学院, 系)^[7]。

5) R 是实体类之间的二元语义关系集(关联关系)。任何所连接的实体类超过两个的关系, 都能够转换为二元的多对一关系集合, 而不丢失任何信息。因此, 该模型中的关系设计为二元关系。 $\forall r \in R$ 可表示为 $N^r(c_1, c_2), c_1, c_2 \in E, N^r$

$\in T$ 为关系的名称, 如 $R(\text{授课}) = (\text{教员, 学员})$ ^[7]。

6) I 实例集, 是信息单元实体类对象集合。

7) F 表示函数集, 主要包括如下函数:

$Att: A \rightarrow E \cup R$, 属性函数, 将属性分配给某一实体类或关系;

$val: V \rightarrow (E, A)$, 属性取值函数, 属性具有数据类型;

$Inst: E \rightarrow 2^I$, 信息单元实例化函数, 可以写为: $Inst(E) = I$ 或 $E(I)$ ^[1];

$Instr: R \rightarrow 2^{I \times I}$, 关系实例化函数, 可以写为: $Instr(R) = \{I_1, I_2\}$ 或 $R(I_1, I_2)$ ^[1]。

8) P 约束规则集, 所有信息单元实体类和关系要满足的约束的集合。包括取值约束和基数约束。

定义 3.3 给定 $MD := \langle S^U, A, H^c, R, L, I, F \rangle$, 语义解释为一个三元组, 记为 $I = \langle \Delta^I, \Delta^L, \cdot^I \rangle$ 。其中 Δ^I 是一个非空集合(包含所讨论领域中的所有个体), Δ^L 是一个非空集合(包含所讨论领域中的所有数据值), \cdot^I 是解释函数, 它将 S^U 中的每个实体类 C 都映射为 Δ^I 的一个子集 $C^I (C^I \subseteq \Delta^I)$, 每个具体域 L 解释成一个集合 $L^I (L^I \subseteq \Delta^L)$, 每个关系 R 解释成一个二元关系 $R^I (R^I \subseteq \Delta^I \times \Delta^I)$, 每个属性 A 解释成一个二元关系 $A^I (A^I \subseteq \Delta^I \times \Delta^L)$, 每个个体 c 解释成一个元素 $c^I (c^I \in \Delta^I)$, 每个值 l 解释成一个元素 $l^I (l^I \in \Delta^L)$ 。

对于每个信息来源来说, 信息源的信息资源可以由多个不同基于本体元数据构成, 不同信息源同样具有不同属性, 因此信息源的描述抽象模型定义如下:

定义 3.4 信息源元数据模版抽象模型 $MSTM := \langle MM, MD \rangle$ 。其中

1) MM 信息资源公共属性描述(描述本元数据所描述的信息源的属性、访问控制等元信息);

2) MD 为定义 3.2 所定义的元数据模型。

4 元数据规范模版

基于图 1 元数据框架层次关系, 我们对于 MST 的设计包括两个层次: 核心层和扩展层。核心层元数据结构的设计是面向共享环境中所有信息资源需要描述的一些共性特征元素的, 通过核心层元数据结构的设计生成相应的核心元数据规范模版(Core Metadata Specification Template, CMST)。扩展元数据结构的设计是面向特定领域、特定任务或特定团体兴趣域的, 它是在核心元数据规范或已有的元数据规范的基础上的扩展, 最终生成领域扩展元数据规范模版(Extensional Domain Metadata Specification Template, EDMST)。因此针对定义 3.2 的元数据模型所定义的实体类、属性、关系等要素的描述所定义的规范模版定义如下:

定义 4.1 实体类的描述 $EM := \langle E_{ID}, C_Meta_{core}, C_Meta_{ext} \rangle$, C_Meta_{core} 为实体类核心描述属性, C_Meta_{ext} 为实体类扩展描述属性, 其中 $C_Meta_{core} := \langle C_title, C_type, C_location, C_subject, C_dynamic, C_importance \rangle$ 。

E (信息单元实体类)是对信息源中一组具有相同属性、关系和语义的对象的描述, 是采用面向对象的方法从信息源中抽象而得到的类。一个信息源通常包含多个实体类。实体类除了具有类所具有的名称、属性等一般特征外, 还扩充了实体类的描述元信息, 以方便进一步的信息处理。 EM 定义实体类的描述元信息, 包括实体类的序列号、名称、类型、位置、描述主题、重要级别、动态系数、视图定义等, 前七项为描述实体类的基本项, 即实体类核心描述属性。

定义 4.2 实体类属性描述 $AM ::= \langle A_Meta_{core}, A_Meta_{ext} \rangle$; 其中 $A_Meta_{core} ::= \langle A_{ID}, A_{title}, A_{type}, A_{location}, A_{subject}, A_{datatype}, A_{key} \rangle$ 。

A (实体类属性)是已被命名了的实体类的特征,它描述了该特性的实例可以取值的范围。 AM 定义实体类属性的元数据描述信息,包括属性的序列号、名称、类型、位置、描述主题、数据类型、关键属性标识、值域、视图定义等,前七项为描述实体类属性的基本项。

定义 4.3 实体类间关系描述 $RM ::= \langle R_{core}, R_{ext} \rangle$, 其中 $R_{core} ::= \langle (E_1, E_2), R_{ID}, R_{title}, R_{type}, R_{location} \rangle$ 。

RM 定义信息类关系的元信息,包括关系所关联的信息类、序列号、存在类型、位置、名称、约束条件等,前六项为描述信息类关系的基本项。关联关系可以具有关系属性,关联关系的属性一般使用与信息类属性相同的元信息描述。关系用于描述信息类与信息类之间的语义连接。信息对象是信息类的实例,信息类与信息类之间的关系代表了信息对象之间的关系。一个信息类可以与它自身存在关联关系,这种关系被称为递归关系。

定义 4.4 信息源元数据描述 $MM ::= \langle MID, Root, S_Meta_{core}, S_Meta_{ext} \rangle$, MID 是MST的标识,用来唯一区分一个MST, $Root$ 是MST的根元素,这是访问MST中信息单元的入口。其中 S_Meta_{core} 定义如下:

$S_Meta_{core} ::= \langle Source_{ID}, Source_{title}, Source_{type}, Source_{location}, Source_{subject} \rangle$ 。

MM 定义信息源的元信息,包括信息源的序列号、名称、类型、位置、信息源内容的描述、动态系数、重要级别、创建记录、使用记录、发布者等。信息源的类型主要包括关系数据库、面向对象数据库、文件系统、XML文件等。相应地,不同类型信息源的位置分别表示不同数据库或文件的实际存在位置,以便最后完成对信息源的检索定位。其中 S_Meta_{core} 为基本项,是描述信息源必须具有的元信息, S_Meta_{ext} 为可选项。

下面对上面定义的几个元数据描述项做一些必要的说明:

- (1) 模型中的元信息项均分为基本项和可选项。其中,基本项是描述必须具备的,其值不能为空,可选项为可选的。
- (2) 可根据应用需要扩充可选元信息,并定义为不同版本的元信息字典。
- (3) 元信息独立于信息源,但不同类型信息源的同项元信息的具体定义规则可能不同。
- (4) 信息类和信息类属性的“类型”、信息类关系的“存在类型”描述它们对应的底部信息源模式内容的具体类型。例如,关系数据库的表或列、XML文件的元素或元素属性,等等。它与“位置”元信息结合使用,以便在查询分解时将子查询映射到物理信息源的相应位置。
- (5) 信息源和信息类的“重要级别”与“动态系数”元信息分别描述信息源与信息类的重要性与动态性。根据它们可以采取相应的策略来进行动态管理。
- (6) 当信息类的“类型”元信息项的值为“视图”时,“视图定义”项才有意义。
- (7) 信息类属性的“关键属性标识”元信息项用于标识该属性是否为所属信息类的关键属性,其取值为“是”或“否”,默认为“否”。

下面我们针对具体核心元数据规范模版进行设计。(CMST)的设计是面向语义信息共享环境中所有信息资源需

要描述的一些共性特征元素的,因而CMST中的元素是所有信息资源元数据描述都必须包含的部分。本节针对信息资源描述、组织、管理和使用的需要对CMST的主要组成进行了设计。

CMST由一个标签为Core_Metadata的复合信息单元进行管理,本文主要以信息源描述元素、网络管理元素为例进行说明。

• 资源描述元素:用于描述或标识信息资源对象内容的外观特征和概要信息的元数据。其结构如图2。

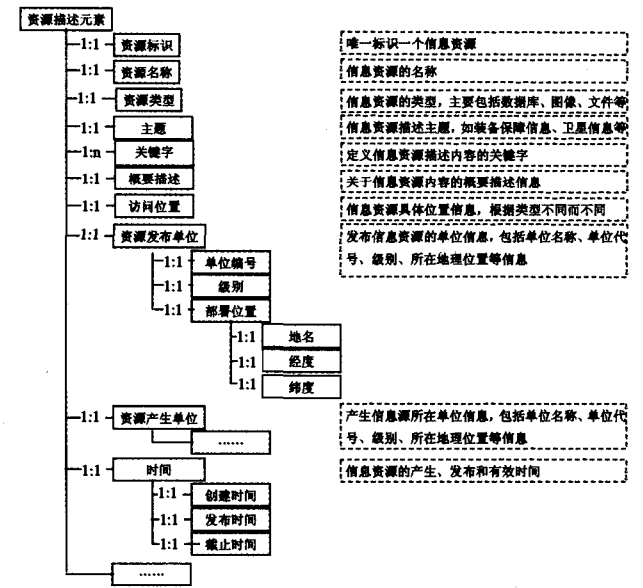


图2 资源描述元素

• 信息资源管理元素:支持信息资源对象网络化管理的元数据。其结构如图3。

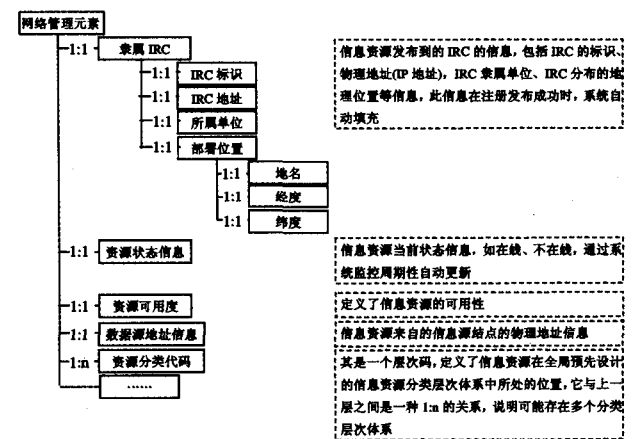


图3 信息资源管理元素

其中资源可用度采用可信度和准确度两项内容加权平均后获得。可信度在资源审核时通过对信息源可信程度和资源合理性判断的基础上赋予一个[0,1]之间的值;准确度在初始状态下赋予一个统一的初值,后面根据用户的反馈评分,周期性地通过平均计算获得,其值也在[0,1]之间。资源分类代码则反映了信息资源在全局分类体系中所属的类别。

5 元数据描述系统原型

本文针对大量异构、多样化的信息资源进行管理的需求,

资源类型包括关系数据库、视频文件、图像文件、已存在的本体对象、Web 网页 (HTML, XML), 以及提供信息的 Web 服务等。为了将这些信息资源有机地融入到本系统的一体化元数据模型中, 需要依据前面介绍的元数据模型对各类信息源中包含元数据和隐含的语义本体知识进行抽取。这一过程主要采取人工注册或半自动抽取的方式, 针对各种类型信息资源建立相应的 wrapper, 基于元数据模型抽取相关元信息和语义内容, 采用 XML 作为元信息交互模型, 生成相应的资源实例对象。

在信息系统中广泛分布的信息资源大量是采用关系数据库的方式进行存储, 关系数据库采用关系模式对信息资源进行结构化存储。关系数据库是关系表的集合, 一般是基于实体联系 (ER) 模型转换而来的, 实体联系 (ER) 模型是对数据库概念建模的重要工具, 是基于世界由一组实体的基本对象及这些对象之间的联系组成, 是一种语义模型。实体联系 (ER) 模型与本体的概念及概念之间的关系相对应, 因此我们可以对应采用本体对信息资源描述, 采用扩展 E-R 图进行可视化描述。

本系统提供了图形化的资源元数据模型封装工具, 如图 4 所示。

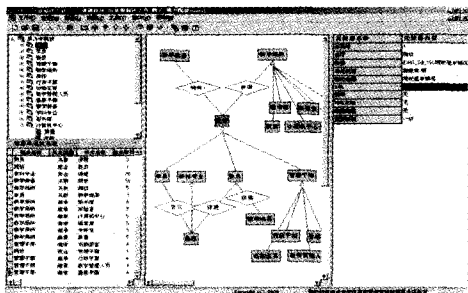


图 4 元数据描述工具运行示例

结论 本文所设计的基于本体的元数据模型可以很好地支持信息资源语义描述, 并通过核心元数据和扩展元数据充分兼顾广域环境下信息资源描述的通用性和特殊性的要求, 并具

有很好的扩展性。元数据模型定义了 IS-A 和 Part-of 的偏序关系以及语义关联关系, 描述了属性以及实例与实体类的映射, 能很好地支持语义相似度的计算, 为基于语义相似度的语义覆盖网构建提供了基础。最后用一个原型系统对元数据模型设计的元数据规范模版加以验证, 证明是切实可行的。

参考文献

- 1 Maedche A. Clustering Ontology-Based Metadata in the Semantic Web
- 2 Wache H, Vogeel T, Visser U. Ontology-Based Integration of Information—A Survey of Existing Approaches. In: Proc. of the UCAI—O1 Workshop on Ontologies and Information Sharing, Seat tie, USA, Aug. 2001. 108~118
- 3 陈汉华, 金海, 等. SemreX: 一种基于语义相似度的 P2P 覆盖网络 [J]. 软件学报, 2006, 17(5): 1170~1181
- 4 刘炜. 元数据与知识本体 [J]. 图书馆杂志, 23(2): 650~54
- 5 王洪伟. 基于本体的元数据模型及 DAM 工表示 [J]. 情报学报, 23(2): 131~13
- 6 宋伟, 张铭. 语义网简明教程. 北京: 高等教育出版社
- 7 刘震, 等. 面向对等网信息语义共享的元数据模型框架研究 [J]. 计算机科学, 2006, 33(1)
- 8 Gruber T R. A translation approach to portable ontologies. Knowledge Acquisition, 1993, 5(2): 199~220
- 9 Wache H, Vogeel T, Visser U. Ontology-based Integration of Information — A Survey of Existing Approaches. In: Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing, Seattle, USA, August 2001. 108~118
- 10 Gruber T. A translation approach to portable ontologies. Knowledge Acquisition, 1993, 5(2): 199~220
- 11 Borst W. Construction of Engineering Ontologies: [Phdthesis]. Enschede: University of Twente, 1997
- 12 Studer R, et al. Knowledge engineering: principles and methods. Data and knowledge engineering, 1998, 25: 161~197
- 13 宋峻峰, 等. OWL DL 的形式化基础研究 [J]. 小型微型计算机, 2005, 26(2): 297~301
- Conference on Performance Modelling and Evaluation of ATM&IP Networks 2001 [C]. Budapest: IFIP, 2001
- 2 Omer G, Israel C, Moshe S. One-way delay estimation using network-wide measurements [J]. IEEE Transactions on Information Theory, 2002, 52(6): 2710~2724
- 3 Bovy C J, Mertodimedjo H T, Hooghiemstra G, et al. Analysis of End-to-End Delay Measurements in Internet [C]. In: Proceedings of the Passive and Active Measurements Workshop (PAM 2002) [C]. Fort Collins (CO), 2002. 26~33
- 4 Zhu Changhua, Pei Changxing. Internet end-to-end delay dynamics [J]. Journal of Systems Engineering and Electronics, 2006, 17(3): 685~691
- 5 Kobayashi K, Katayama T. Analysis and Evaluation of Packet Delay Variance in the Internet [J]. IEICE Trans Commun, 2004, E85-B: 35~41
- 6 焦利, 林宇, 王文东, 等. 一种负载均衡网络中内部链路时延推测算法 [J]. 软件学报, 2005, 14(5): 886~893
- 7 Moon S B, Skelly P, Towsley D. Estimation and Removal of Clock Skew from Network Delay Measurements [A]. In: IEEE Infocom99 [C]. New York: IEEE, 1999. 227~234

(上接第 111 页)

实例 2 源主机 (西电校园网), 目的主机 (西安交大校园网)

测试条件: 发送 1000 个 500 字节长数据包, 时间间隔为 50ms, 从源主机 (西电校园网) 到目的主机 (西安交大校园网), 同为 CERNET 网, 共 6 跳。同样也可以看出, 由于时钟不同步的影响, 端到端时延实测结果呈现一个线性趋势; 消除时钟不同步带来的影响后, 时延变化趋于平稳。

结论 本文给出了测量网络端到端时延的方法, 根据测得的结果, 用线性规划的方法来消除收发时钟的初始相位差和相对频差等参数, 获得端到端时延的真实值。实例测量结果表明, 该方法消除了时钟不同步带来的影响, 提高了测试结果的精确度, 可以应用于互联网中的端到端时延测量。

参考文献

- 1 Van Moffaert A, De Vleeschauwer D, Janssen J. Tuning the VoIP Gateways to Transport International Voice Calls over a Best-Effort IP Backbone [A]. In: Proceedings of the 9th IFIP