

基于多本体语义相似度计算的对等网资源 动态匹配方法研究*)

刘震 邓苏 罗雪山 黄宏斌

(国防科技大学信息系统与管理学院 长沙 410073)

摘要 在基于对等网的网格环境中,应用本体来描述资源元数据已经成为当今研究的趋势。不同结点可能采用不同的本体来描述各自管理的资源元数据。如何提高这种环境下动态资源发现的效率是急需研究的问题。本文在形式化描述结点本体结构和元数据模型结构的基础上,提出了一种精确匹配与模糊匹配相结合、本体语义相似性匹配和资源实例特征匹配相结合的综合资源匹配方法,来提高非集中环境下资源匹配的灵活性和精确性。

关键词 网格,对等网,本体,元数据,资源匹配

Research on Dynamic Resource Matching Approach Based on Multi-Ontology Semantic Similarity in Peer-to-Peer Network

LIU Zhen Deng Su LUO Xue-Shan HUANG Hong-Bin

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

Abstract In peer-to-peer based grid environment, applying ontology to describe resource has become current research trend. However, different peer may use different ontology to describe resource metadata according to its domain knowledge. How to matching these resources is a desiderate research problem in this environment. In this paper, based on formal description of peer ontology and metadata model, a hybrid resource matching approach is proposed. It combines precision matching and fuzzy matching. At the same time, it combines semantic matching and resource instance characteristics matching. This approach will improve flexibility and accuracy of resource matching in decentralized environment.

Keywords Grid, Peer-to-peer, Ontology, Metadata, Resource matching

1 引言

网格与对等网技术之间具有很强的互补性,两者的融合^[1]也已经成为学术界的共识。随着网格与对等网的融合,网格资源具有以下特点:1)规模大、分布、无集中控制。2)异构性和多样性。随着网格应用领域的扩张,如军事、教育、电子政务等领域,管理资源的类型和特性差异将急剧增大。可能包括的资源类型:a)异构数据资源,包括结构化、半结构化和无结构数据,如关系数据库、XML、多媒体数据、文件等;b)服务资源,如Web服务或远程过程等;c)物理资源描述信息,如处理器类型、存储能力等计算和存储资源的描述信息。d)集成资源,如对多个数据源的集成视图。

面向网格中资源的新特点,基于传统的简单资源标识匹配或基于属性的简单资源特性描述匹配,将无法充分满足用户或应用面向资源语义内容的资源发现需求。应用本体技术来描述网格资源元数据已经成为当前研究的一个热点。然而在一个大规模、非集中的网络环境中,采用一个全局一致本体来进行描述,是不切实际的。因而必然存在多个本体对类似资源进行描述,那么在这种多本体环境中,如何提供一种动态、有效的资源匹配方法是一个亟待研究的问题。

本文面向基于对等网的信息网格环境下资源发现问题,在形式化描述基于本体的结元数据模型结构的基础上,提出了一种精确匹配和模糊匹配相结合、基于本体的语义相似

度计算和资源实例特征相似度计算相结合的一种综合的动态资源匹配方法。第2节首先对相关研究工作介绍;第3节对资源组织结构和基于本体的资源元数据描述模型进行介绍;第4节对基于语义的综合资源匹配方法需要考虑的要素及匹配过程进行了研究设计。第5节结合教育领域的一个资源匹配实例,对匹配过程进行了说明。

2 研究现状及相关工作

目前网格系统中资源发现基本上是集中式的,在Globus中MDS实现了基于LDAP的树状元数据目录服务来支持资源发现;在Condord的matchmaker^[2]实现了不依赖全局资源命名,而依靠属性匹配的集中式资源发现机制。Web服务中的UDDI^[3]实现了集中式的服务实体的统一描述、注册和查找。在这些系统中,主要采用的是一种精确匹配的机制,然而其资源描述机制的局限,限制了其提供的资源发现能力。

在对等网领域,如Gnutella^[4], Napster^[5], CAN^[6], Pstry^[7]等,往往采用关键字或全局标识(如文件名)来作为资源元数据描述,查询主要采用字符串匹配的方式,大大地局限了资源匹配的灵活性和面向内容的资源定位能力。EDUTEL-LA^[8]、SWAP^[9]是基于对等网体系结构的知识共享研究项目,它们都采用RDF来描述元数据,并分别提供相应的查询语言,如RQL, SeRQL,来提供面向语义内容的查询,它们还没有对异构本体下的资源匹配问题做深入考虑。Edamok^[13]

*)国家自然科学基金(批准号:60172012)。刘震 博士研究生,主要研究领域为信息管理、P2P、智能决策支持技术;邓苏 教授,主要研究领域信息管理、智能决策支持技术;罗雪山 教授、博士生导师,主要研究方向为指挥自动化。

也是面向对等网环境下知识共享的研究项目,为了发现存储在不同结点上概念间的映射关系,提供了 CTX-MATCH 算法,它主要采用描述逻辑的方法。HELIOS^[11]也是面向对等网的本体知识共享的项目,其中采用 H-Match 算法来支持不同本体概念之间语义匹配度的动态计算,但没有考虑资源实例特征因素。本文的特点在于提供了一种综合的资源匹配方法,其中结合了精确匹配与模糊匹配特点,并且综合考虑了本体语义相似度和资源实例特征相似度,提供了更为灵活的、高效的资源匹配能力。

3 资源组织结构及元数据模型

3.1 资源组织结构

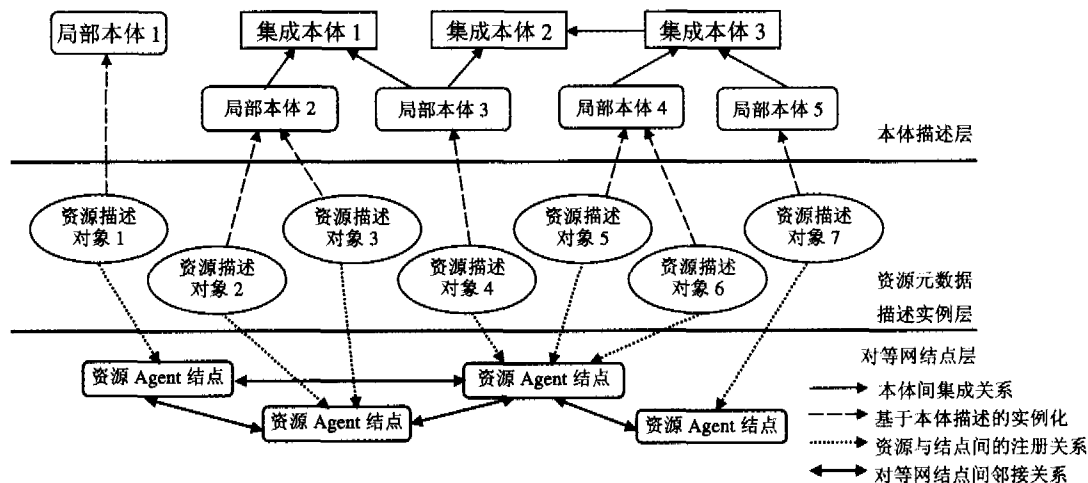


图1 基于复合多本体结构的资源组织结构

3.2 基于本体的元数据描述模型

本体从不同的问题领域和研究角度出发存在许多不同的定义。Studer 等在对本体进行了深入研究之后,给出了一个广为接受的定义,即:“本体是共享概念模型的明确的形式化规范说明”^[10]。本文中资源元数据采用一种基于本体的模式进行描述。

本文中考虑要描述元数据包括两个层次:核心层(Core Layer)和扩展层(Extensional Layer)。核心层是系统应用环境内每一个资源描述都包含的部分,对资源的整体特征进行描述,其中包括资源管理元素、概要元素、物理特征元素、安全和信任元素。扩展层则是提供面向特定领域的或特定任务的领域本体,是对核心层的扩展,提供粒度层次更高的、面向语义内容的资源描述,如针对数据库资源的模式语义的描述,多媒体信息语义内容的抽取等,将有效地提高信息资源检索的效率和精度。

定义1 结点本体结构 $O_{node} = (C, A, R, P)$ 。T 是术语集合, D_{basic} 是预定义的基本数据类型集合。

$C = C_{key} \cup C_{ext}$ 是结点中所有领域概念(Concept)的集合,其中 C_{key} 为核心层概念集合, C_{ext} 为扩展层概念集合。 $C_{ext} = C_{loc} \cup C_{net}$, 其中 C_{loc} 为本地资源概念集合, C_{net} 为远程资源概念集合。一个概念的可以形式化定义为 $c = (t_c, A_c)$, 其中 $c \in C, t_c \in T, A_c \subseteq A$, 表示 c 的属性特征集合。

A 是所有属性的集合。每个属性可以形式化定义为 $a = (t_a, dt_a)$ 。 $t_a \in T, dt_a \in D_{basic} \cup C_{loc}$ 。

$R = R_{loc} \cup R_{net}$ 定义了概念之间的二元语义关系集合,其中 R_{loc} 定义了本地概念关系的集合, $\forall r_{loc} \in R_{loc}$ 表示为 $r_{loc} \langle c,$

本文采用对等资源 Agent 网络来管理多个组织的分布异构的各类资源。资源的组织管理结构层次如图1所示,每个资源 Agent 结点依附于一个具体单位或一个虚拟组织,管理着注册的一个或多个资源对象。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,采用基于本体的方法来描述资源的元数据,面向对等网的分布特征,采用一种复合多本体结构来描述和组织管理分布异构资源的元数据,其中每个本体都是建立在某网络结点所在单位对自己相关领域的理解基础上。多个本体可能由于领域背景不同而在类似概念描述上存在差异。如何在这种不同本体环境下有效地匹配到类似的相关资源是本文要研究的工作。

$c', c, c' \in C_{key} \cup C_{loc}$ 。 $\forall r_{net} \in R_{net}$ 表示为 $r_{net} \langle c, c' \rangle, c \in C_{loc}, c' \in C_{net}$ 。本系统中描述的语义关系主要包括:等价关系 R^{same} 、继承关系 H^{inh} 、聚合关系 H^{agr} 、关联关系 R^p 。

P 是 O_{node} 中的约束集合。 P_{key} 是关键字约束集合, $\forall p_{key} \in P_{key}$, 可形式化表示为 $p_{key} = (c, A_{key}), c \in C, A_{key} \subseteq A_c$ 。 P_{attr} 是属性约束集合, $\forall p_a \in P_{attr}$, 可形式化表示为 $p_a = (a, c, k), a \in A, c \in C, k \in \{0, 1\}, k=0$ 表示 a 是 c 的可选属性, $k=1$ 表示 a 是 c 的强制属性。

定义2 结点元数据模型结构 $M_{node} = (O_{node}, S, I, IF_s^c, IF_s^i, IC)$ 。

O_{node} 是元数据模型参考的结点本体。

S 是结点中管理的所有资源对象集合,每个资源对象对应一个资源提供者,由一个资源标识唯一确定。

I 是结点内所有概念实例集合。 $I(s, C)$ 表示资源 s 针对概念集合 C 的实例数据集合。

IF_s^c 是所有资源到对应概念实例映射函数集合,其中 $\forall f \in IF_s^c$, 可表示为 $f: s \times c \rightarrow I'$, 其中 $s \in S, c \in C, I' \subseteq I$ 。 $IF_s^c = IF_{static}(S, C) \cup IF_{dyn}(S, C)$, 其中 $IF_{static}(S, C)$ 对应静态映射函数集合, $\forall if_{static} \in IF_{static}(S, C)$, 可表示为 $if_{static}(s, c) = \{ \langle \langle a_1, v_{11} \rangle, \dots, \langle a_m, v_{1m} \rangle \rangle, \dots, \langle \langle a_1, v_{n1} \rangle, \dots, \langle a_m, v_{nm} \rangle \rangle \}$, 其中 $s \in S, c \in C, \{a_1, \dots, a_m\} \subseteq A_c, v_{11}, \dots, v_{1m}, \dots, v_{n1}, \dots, v_{nm}$ 是一组常量值,即是说对实例数据与元数据存储在一起。 $IF_{dyn}(S, C)$ 对应动态映射函数集合。采用链接或松耦合的逻辑规则(Logical Rules)和公理(Axioms)来定义这种关系,即实例数据采取虚拟存储的形式,要通过与底层模式链接或推理计算后才能获得。

IF_S 是所有资源到对应概念关系实例映射函数集合, 其中 $\forall f \in IF_S$, 可表示为 $f: s \times r \rightarrow I' \times I''$, 其中 $s \in S, r \in R_{loc}, I', I'' \subseteq I$.

IC 是资源对象实例特征约束集合。对于任意约束 $ic \in IC$ 用一个规一分布来表示, $ic(s, C_k, A_k) = \{ \langle pic_1, p_1 \rangle, \langle pic_2, p_2 \rangle, \dots, \langle pic_n, p_n \rangle \}$, 其中 $s \in S, C_k \subseteq C(s), A_k \subseteq A(C_k), \bigcup_{i=1}^n I(s, C_k, A_k, pic_i) = I(s, C_k, A_k), \sum_{i=1}^n p_i = 1, p_i \geq 0, C(s)$ 为资源 s 实例化的本体 O_{node} 中的概念子集。 $I(s, C_k, A_k, pic_i)$ 表示 s 中满足约束 pic_i 的 C_k 的属性 A_k 实例集合, p_i 体现了 $I(s, C_k, A_k)$ 满足约束 pic_i 的可能性, $p_i = 1$ 时表示 $I(s, C_k)$ 必须满足约束 pic_i 。这种实例约束特征可以由专家定义或通过数据挖掘方法自动发现和抽取。

4 基于语义相似度匹配的资源发现方法

为了增强非集中环境下动态资源发现和查询的能力, 需要一种灵活有效的模糊匹配算法来发现网络中具有回答相关查询能力的资源。下面将基于前面的本体及元数据模型描述框架, 结合核心概念特征、术语特征、语义环境特征和实例特征这几个方面来设计相应的匹配算法。

4.1 查询模型

用户或相关应用的资源查询请求描述往往依据依附的结点本体进行定义。查询的目标就是为了获得一组用户关注的资源信息实例。结点本体结构可以看成由一组概念和关系构成的语义有向连通图, 则查询往往涉及这个语义图中的一个子图。

定义3 查询模型 $Q = (C_Q, R_Q, qc)$, 其中 C_Q 和 R_Q 的结构定义与前面结点本体 O_{node} 中 C_Q 和 R_Q 的定义相同, C_Q 表示查询中涉及的概念集合, R_Q 表示查询中涉及的关系集合。 qc 定义了查询中概念实例必须满足的约束条件, 可表示为 $qc(C_q) = qc_1(c_{q1}, c_{q2}, \dots, c_{qn}) \vee qc_2(c_{q1}, c_{q2}, \dots, c_{qn}) \vee \dots \vee qc_n(c_{q1}, c_{q2}, \dots, c_{qn}), C_q \subseteq C_Q, C_q = \{c_{q1}, c_{q2}, \dots, c_{qn}\}, qc_i(c_{q1}, c_{q2}, \dots, c_{qn})$ 表示针对 $\{c_{q1}, c_{q2}, \dots, c_{qn}\}$ 的某个子集的相关条件约束的合取表达式, 即原子条件约束。

查询 Q 可依据查询约束分解为一系列的原子查询 $Q_1, Q_2, \dots, Q_n, Q_i = (C_Q, R_Q, qc_i)$ 。

本文研究的资源匹配算法的目标就是根据查询需求发现具有应答能力的备选资源元数据, 并提供相应资源的匹配度。

4.2 匹配算法

4.2.1 基本匹配要素

• 核心层概念的匹配。核心层概念是在所有结点本体中都包含的部分, 对于核心层概念的匹配采取实例级的精确匹配。查询与结点某资源在核心概念 c_k 上的匹配度 $KA(Q, s, c_k)$ 的计算方法如下:

$$KCSim(s, C_k, qc_i) = \begin{cases} 1, & \text{如果 } \exists qc_i(c_k), \text{ 且 } I(s, C_k) \cap I(s, C_k, qc_i(c_k)) \neq \emptyset \\ 1, & \text{如果不存在 } qc_i(c_k) \\ 0, & \text{否则} \end{cases}$$

其中 $s \in S, c_k \in C_q, c_k \in C_{Key}, qc_i(c_k)$ 表示原子条件约束中涉及 c_k 的部分。

• 扩展层概念的匹配。扩展层概念在不同结点具有一定的差异。由于用户视角或领域的不同, 可能对类似资源信息采用不同的概念描述。针对扩展概念的匹配, 本文采用类似 H-Match^[11] 的方法来进行匹配。从概念描述的术语特征和语义环境特征两个方面进行匹配。

a) 命名特征。每个概念和属性描述都包含一个名字(词汇和术语), 来较贴切地描述其内在的涵义。本文考虑利用电子字典 WordNet^[12] 来发现术语间关系, 术语关系包括: 同义、狭义、广义、相关。对于术语间的不同关系赋予一个权值来表达语义近似程度, 表示为 $W_{syn}, W_{nar}, W_{gen}, W_{ass}$, 且 $W_{syn} > W_{nar} \geq W_{gen} > W_{ass}$, 取值在 $[0, 1]$ 。术语相似度计算方法如下:

$$TermSim(t, t') = \begin{cases} \text{Max}(W_{t1} * W_{t1t2} * \dots * W_{tnt'}), & \text{如果存在 } t \rightarrow t' \\ 0, & \text{否则} \end{cases}$$

其中 $t \rightarrow t'$ 表示存在从术语 t 到 t' 的术语关系路径, W_{tij} 表示术语 t_i 和 t_j 之间的关系的权重, 可能对应 $W_{syn}, W_{nar}, W_{gen}, W_{ass}$ 中的一种。

定义4 概念间命名相似度 $NameSim(c, c') = Termsim(t_c, t_{c'})$ 。

b) 语义环境特征。概念间的语义相似性除了从命名特征考虑外, 还要考虑概念语义环境相似性。

定义5 对于任意概念 $c \in C_{loc}$, 其语义环境 $SEvm(c) = A(c) \cup RC(c)$, 其中 $A(c) = \{a_i | a_i \in A_c\}, RC(c) = \{c_j | r(c, c_j), r \in R_{loc}\}$ 。

对于 $\forall e \in SEvm(c), \forall e' \in SEvm(c'), e, e'$ 之间的匹配度计算函数如下:

$$ElemMatch(e, e') = \begin{cases} TermSim(n_e, n_{e'}) * dtSim(dt_e, dt_{e'}) * (1 - |W_e - W_{e'}|) & \text{如果 } e \in A(c), e' \in A(c') \\ TermSim(n_e, n_{e'}) * (1 - |W_e - W_{e'}|), & \text{否则} \end{cases}$$

其中 $n_e, n_{e'}$ 是环境因子 e, e' 的命名术语, $dtSim(dt_e, dt_{e'})$ 定义了预定义数据类型之间的兼容度, 取值在 $[0, 1]$ 。 $W_e, W_{e'}$ 分别对应 c 与 e 之间关系、 c' 与 e' 之间关系的影响因子权重。

定义6 概念 c, c' 之间语义环境相似度计算方法如下:

$$EvmSim(c, c') = \frac{\sum_{i=1}^n \max((ElemMatch(e_i, e')) | e' \in SEvm(c'))}{|SEvm(c)|}$$

c) 语义匹配度。概念 c, c' 之间的语义匹配度为:

$$SemanticMatch(c, c') = W_{NS} * NameSim(c, c') + (1 - W_{NS}) * EvmSim(c, c')$$

• 实例特征匹配。实例特征匹配考虑通过元数据模型描述中实例约束和相关知识特征来评估资源概念实例与查询结果实例之间的可能匹配度, 也就是从实例层评估相关资源回答查询的能力。

定义7 概念实例特征匹配度计算函数表示如下:

$$InsMatch(c, c', s) = \begin{cases} 0, & \text{如果不存在 } if(s, c') \in IF_S \text{ 即 } I(s, c') = \emptyset \\ 1, & \text{否则如果不存在 } qc_i(c) \\ 1, & \text{否则如果 } if(s, c') \in IF_{static} \text{ 且 } \exists qc_i(c), \\ & \text{且 } I(s, c') \cap I(s, c', qc_i(c')) \neq \emptyset \\ a, & \text{否则如果 } \exists qc_i(c) \text{ 且不存在 } ic(s, c') \\ & \text{或 } A(qc_i(c)) \cap A(ic(s, c')) = \emptyset \\ \prod_{i=1}^k (\sum_{j=1}^m p_{ij} * b_{ij}), & \text{否则如果 } if(s, c') \in IF_{static} \\ & \text{且 } \exists IC(s, c') = \{ic_1, \dots, ic_k\} \neq \emptyset, \text{ 其中 } \forall ic_i \in IC(s, c'), \\ & ic_i(s, c') = \{ \langle pic_{i1}, p_{i1} \rangle, \dots, \langle pic_{im}, p_{im} \rangle \}, \\ & \text{且满足 } A(qc_i(c')) \cap A(ic(s, c')) \neq \emptyset \\ & \text{对于 } \forall \langle pic_{ij}, p_{ij} \rangle \in ic_i(s, c'), \\ & \text{当 } I(s, c', pic_{ij}) \cap I(s, c', qc_i(c')) \neq \emptyset \\ & \text{则 } b_{ij} = 1, \text{ 否则 } b_{ij} = 0 \end{cases}$$

其中 $s \in S, c' \in C(s), c \in C_Q, \alpha$ 为一个系统预定义常量, 取值在 $[0, 1]$ 中, 作为默认的实例匹配度。 $IC(s, c')$ 表示 s 中所有涉及 c' 的实例约束集合。 $A(*)$ 表示 $*$ 中涉及的属性集合。

4.2.2 资源匹配方法过程

前面已经提到, 查询 Q 可依据查询约束分解为一系列的原子查询 $Q_1, Q_2, \dots, Q_n, Q_i = (C_Q, R_Q, q_i)$ 。本文描述综合资源匹配方法主要是针对原子查询 Q_i 进行, 最后将原子查询匹配结果进行合并。过程描述如下:

输入: 原子查询 $Q_i = (C_Q, R_Q, q_i)$, 网络结点 i 的元数据知识库模型 $M_{nodei} = (O_{node}, S, I, IF_S^C, IF_S^R, IC), \lambda_1, \lambda_2, \lambda_3, \alpha, W_{NS}$

输出: 资源匹配状态集合 $SS_Q = \{(s_1, sm_1), \dots, (s_m, sm_m)\}$

开始:

Step1 初始 $SS_Q = \phi$, 提取 C_Q 中涉及核心层概念 C_{Qk} , 若不存在 $q_i(C_{Qk})$, 跳到第 2 步执行。若 $\exists q_i(C_{Qk})$, 对于 $\forall s \in S, \forall c_k \in C_{Qk}$, 将使得 $KCSim(s, c_k, q_i) = 1$ 的 $(s, 1)$ 加入 SS_Q , 若没有匹配的资源 s , 则算法退出。否则继续第 2 步执行。

Step2 提取 C_Q 中涉及扩展层概念 C_{Qe} 。如果 $C_{Qe} = \phi$, 算法结束, 否则继续。如果 $SS_Q \neq \phi$, 获得 SS_Q 中的资源集合 $S', C_S = C(S') \cap C_{ext} \cap C_{loc}$; 如果 $SS_Q = \phi, S' = S, C_S = C_{ext} \cap$

C_{loc} 。

Step3 计算扩展层概念语义匹配度。对于 $\forall c \in C_{Qe}, \forall c' \in C_S$, 首先计算 $NameSim(c, c')$, 针对 $NameSim(c, c') \geq \lambda_1$ 的概念匹配对, 计算 $EvmSim(c, c')$, 并最终计算获得 $SemanticMatch(c, c')$ 。 $CPair = \{(c, c') | NameSim(c, c') \geq \lambda_1, SemanticMatch(c, c') > \lambda_2, c \in C_Q, c' \in C_S\}$ 表示概念语义匹配对集合, $C_{QMatch} = \{c | (c, c') \in CPair\}$ 表示查询中获得匹配的概念集合。

Step4 计算实例特征匹配度。对于 $\forall (c, c') \in CPair$, 且 $\forall s \in S', \forall c' \in C(s)$, 分别计算 $InsMatch(c, c', s)$ 。

Step5 计算资源匹配度。对于 $\forall s \in S'$, 计算资源匹配度 $ResourceMatch(s, Q)$, 计算方法如下:

$$ResourceMatch(s, Q) = \frac{\sum_{c \in C_Q} \text{Max}(\{SemanticMatch(c, c') * InsMatch(c, c', s) | c' \in C(s)\})}{|C_Q| - |C_{Qk}|}$$

将 $ResourceMatch(s, Q) \geq \lambda_3$ 的 $(s, ResourceMatch(s, Q))$ 加入 SS_Q , 若 SS_Q 中已存在 $(s, 1)$, 用 $(s, ResourceMatch(s, Q))$ 覆盖加入。

Step6 结束。

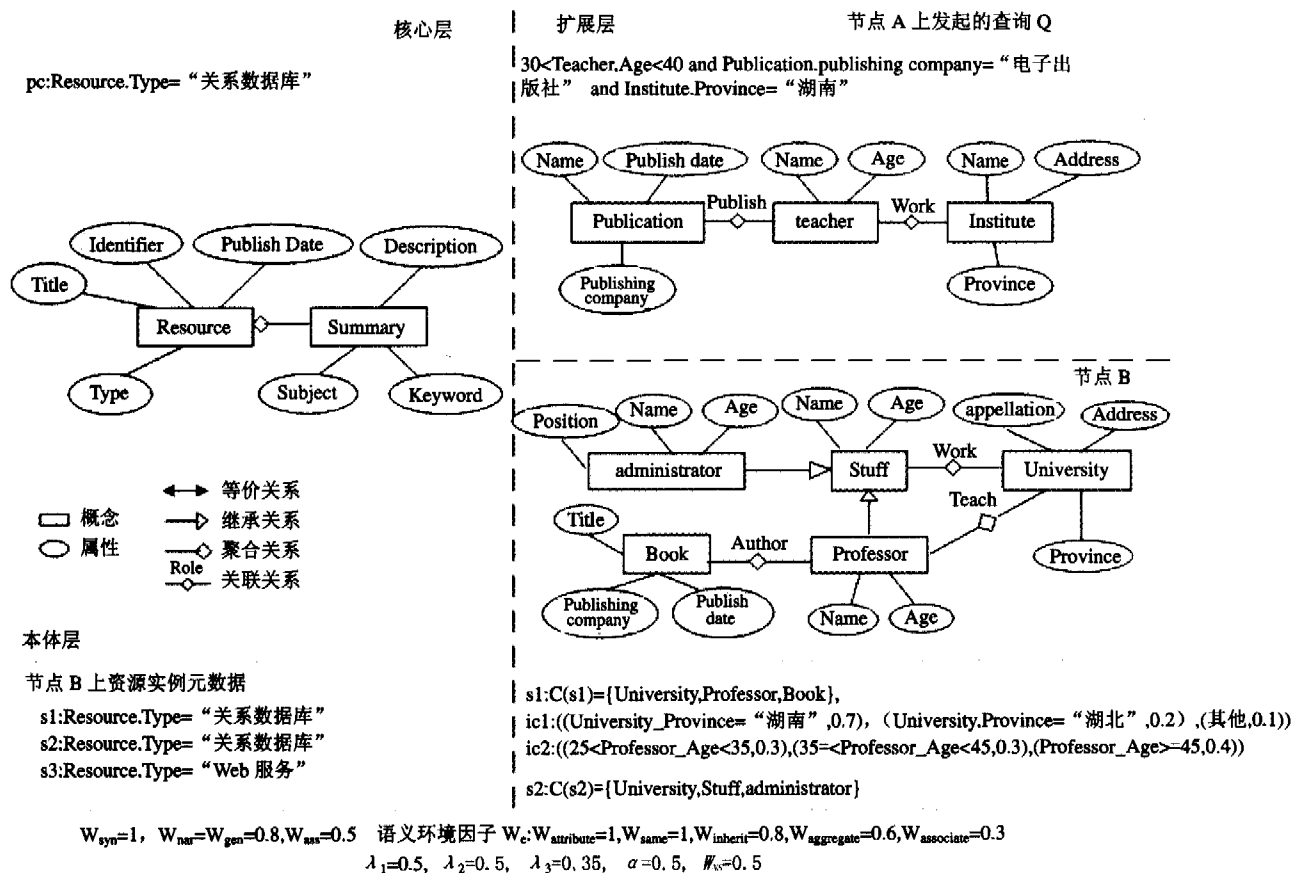


图 2 结点元数据及查询描述示例

4.3 应用示例

下面将通过一个简单示例来描述资源匹配算法。图 2 给出了一个教育领域的结点本体、元数据及查询示例。结点 A、B 分别对应两个院校。结点 B 中管理着三个资源 s_1, s_2, s_3 。结点 A 依据本地本体并针对全局所有院校发起的查询 Q , 当查询传递到结点 B, 要发现结点 B 中具有应答查询能力的所

有资源, 因而依据 4.2.2 节中描述的资源匹配过程, 首先进行核心层概念的精确匹配, 可以获得满足 $Resource.Type=“关系数据库”$ 的资源 s_1, s_2 , 因而可以排除 s_3 资源; 接着进行扩展层概念语义匹配, 得到满足阈值 $\lambda_2=0.5$ 的概念匹配对的语义匹配度 $SemanticMatch(Institute, University)=0.8, SemanticMatch(teacher, stuff)=0.65, SemanticMatch(teacher, ad-$

ministrator) = 0.53, SemanticMatch (teacher, professor) = 0.75, SemanticMatch(publication, book) = 0.85。第三步计算资源实例匹配度, 针对 s_1 的实例特征匹配度 $\text{InsMatch}(\text{Institute}, \text{University}, s_1) = 0.7$, $\text{InsMatch}(\text{teacher}, \text{professor}, s_1) = 0.3 + 0.3 = 0.6$, $\text{InsMatch}(\text{publication}, \text{book}, s_1) = 0.5$ 。由于 s_2 中没有定义实例特征约束, 因而 s_2 中所有实例特征匹配度均为 $\alpha = 0.5$ 。最后计算资源元数据匹配度, $\text{ResourceMatch}(s_1, Q) = (0.8 * 0.7 + 0.75 * 0.6 + 0.85 * 0.5) / 3 = 0.48$, $\text{ResourceMatch}(s_2, Q) = 0.29$ 。于是根据阈值, $\lambda_3 = 0.35$ 得到匹配的资源 s_1 , 并将 s_1 的相关元数据和匹配结果(可以包括概念匹配关系等)发送到结点 A。

资源匹配过程可以通过改变参数 $\lambda_1, \lambda_2, \lambda_3, \alpha, W_{NS}$ 来调节匹配的性能。

结束语 本文针对基于对等网的网格环境中的资源发现问题, 在形式化描述基于本体的结点元数据模型结构的基础上, 提出了一种基于核心元数据的精确匹配和基于扩展元数据的模糊匹配相结合、基于本体语义相似度匹配和资源实例特征匹配相结合的综合资源匹配方法, 将有效提高网格资源匹配的精度和灵活性。

资源请求路由策略也是资源发现系统需要考虑的一个核心问题, 下一步工作将提供一种高效的资源请求路由策略。在基于对等网结构的资源管理模式, 通过发现结点本体间语义关联性, 将相关结点聚类在一起, 形成社区或构建超级结点, 来自发请求的传播路由, 从而减少对等网环境下查询请求传播的盲目性, 提高资源定位的效率。

(上接第31页)

时呼叫的平均等待时间较小, P_{loss} 也较低, 能充分有效地利用网络资源。此时系统的 QoS 较高。表 1 为该模拟移动系统当 $\rho_1 = 0.85, T_1 = 15\%W_s, T_2 = 3 \sim 5 T_1$ 时的拥塞控制部分仿真结果。其中, 结果 1 和结果 2 为网络在处理 1000 个呼叫的时候, 网络一直没有出现拥塞情况, 这种现象一般发生在呼叫低谷期, 例如每天凌晨时段, 呼损率均为 0; 结果 3、4、5 的平均缓解拥塞的时间大约为 $15\%W_s$, 可知该种情况下, 移动系统时有拥塞发生, 但皆为一般拥塞状态。启动时钟 T_1 , 拥塞得到及时缓解, 且此时呼损率几乎为零, 呼叫平均等待时间接近 W_s ; 结果 6、7 显示了当网络发生一般拥塞和严重拥塞两种情况时系统的运行情况, 尽管以拥塞缓解时间较长(为 $40\%W_s$ 左右)、呼叫延时较大、呼损率上升等为代价, 但系统出现严重拥塞状态时, 能够得到及时处理, 保证了系统充分地利用资源, 能够稳定工作。

表 1 拥塞控制程序运行结果

	平均等待时间	拥塞缓解时间	呼损率
结果 1	14.2331	0.0000	0.00
结果 2	15.7216	0.0000	0.00
结果 3	18.2866	2.4604	0.00
结果 4	18.3616	2.6649	0.00
结果 5	17.3985	2.5607	0.00
结果 6	35.7651	13.7339	5×10^{-4}
结果 7	38.160	14.301	5×10^{-4}

6 融于速率控制策略的蜂窝移动通信系统的拥塞“软处理”方法的特点

1. 在蜂窝移动通信系统中, 根据上行链路 SIR 与 SIR_{eq}

参考文献

- 1 Talia D, Trunfio P. IEEE Internet Computing: Toward a Synergy Between P2P and Grids. IEEE Distributed Systems Online 2003, 4(7):3
- 2 Raman R, Livny M, Solomon M. Matchmaking: Distributed Resource management for high throughput computing. In: Proc. of the 7th IEEE HPDC. Washington, DC: IEEE Computer Society Press, 1998. 140~146
- 3 UDDI.org, 2000. <http://www.uddi.org>
- 4 Ripeanu M, Foster I, Iamnitchi A. Mapping the Gnutella network, Properties of large-scale peer-to-peer systems and implications for system design. Internet Computing 2002, 6(1):50~57
- 5 <http://www.napster.com>
- 6 Ratnasamy S, Francis P, Handley M. A Scalable Content-Addressable Network. In: Proceedings of ACM SIGCOMM, 2001
- 7 Rowstron A, Druschel P. Pastry: Scalable, Distributed Object Location and Routing for Large-scale Peer-to-Peer Systems. In: IF-IP/ACM International Conference on Distributed Systems Platforms(Middleware), Heidelberg, Germany, 2001. 329~350
- 8 Nejd W, Wolf B, Qu C, et al. A P2P networking infrastructure based on rdf. In: Proceedings to the Eleventh International World Wide Web Conference, Honolulu, Hawaii, USA, 2002
- 9 Ehrig M, Tempich C, Broekstra J, et al. SWAP - ontology-based knowledge management with peer-to-peer technology. Sure Y, Schnurr, H. P(eds) Proceedings of the 1st National Workshop
- 10 Studer R, Benjamins V R, Fensel D. Knowledge Engineering, Principles and Methods. Data and Engineering, 1998, 25(1-2):161~197
- 11 Castano S, Ferrara A, Montanelli S. H-MATCH: an Algorithm for Dynamically Matching Ontologies in Peer-based Systems. In: Proc. of the 1st Int Workshop on Semantic Web and Databases (SWDB) at VLDB, 2003, Berlin, Germany, September 2003
- 12 Miller A. WordNet: A lexical database for english. Communications of the ACM, 1995, 38(11):39~41
- 13 Bouquet P, Magnini B, Serafini L, et al. A SAT-based algorithm for context matching. In: Proceedings of the 4th International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT-03), Springer Verlag, 2003. 66~79
- 14 Tangmunarunkit H, Decker S, Kesselman C. Ontology-based resource matching in the grid - the grid meets the semantic web. Proceedings of SemPGRID '03, Budapest, May 2003

比较结果, 确定移动网络的运行情况, 对呼叫采取不同的处理。当系统发生拥塞时, 通过拦截优先级较低的呼叫, 同时启动控制时钟, 实现对移动通信网络拥塞缓解的目的。

2. 通过设定两个时钟 T_1 和 T_2 , 使得移动网络的拥塞及时得到缓解, 特别是当移动网络某局部出现严重拥塞时, 时钟 T_2 的设定为尽快解除拥塞起了关键作用。可见, 该控制方法具有一定的完善性。

3. 该控制模型作用于基站, 能够保证缓解拥塞时间短、呼叫等待时间短、呼损率低的特点。特别在高峰期, 能有效地、及时地解除拥塞, 具有现实意义。

4. 该控制方法将预防式控制与反应式控制方法相结合, 运用起来简单, 易于实现。

5. 由于对呼叫的到达率采用控制机制, 具有更短的队列长度、更短的等待时间和较小的拥塞缓解时间, 有效地提高了移动通信网络的服务质量。

参考文献

- 1 Liu Zhao, El Zarki M. SIR-Based Call Control for DS-CDMA Cellular Systems [J]. IEEE J Select Areas in Commun, 1994, 12: 638~644
- 2 Zander J. Performance of optimal Power Control in Cellular Radio Systems [J]. IEEE Trans Ved Tech, 1992(1)
- 3 Zander J. Radio Resource Management in 3rd Generation Personal Communications Systems [J]. IEEE Comm Mag, 1998(8)
- 4 Chen Xilin. Modern Communications Theory [D]. EI Publish Company of Bei jing, 1999
- 5 Gelenbe E, Pujolle G. Introduction to queueing Networks [D]. Editions Frvolles, 1987
- 6 Werner R P, Konhauser M W. Third Generation Mobile Communication Systems [J]. Publishing House of Electronics Industry, 2001, 1:169