

互网络中语义媒体查询的关系路由方法*

廖祝华^{1,2,3+}, 张国清¹, 杨景^{1,4}, 傅川¹, 张国强^{1,5}

¹(中国科学院 计算技术研究所, 北京 100190)

²(中国科学院 研究生院, 北京 100049)

³(知识处理和网络化制造湖南省高校重点实验室(湖南科技大学), 湖南 湘潭 411201)

⁴(中国移动通信研究院, 北京 100053)

⁵(南京师范大学 计算机科学与技术学院, 江苏 南京 210012)

Scheme of Relational Routing for Querying Semantic Media on Internetwork

LIAO Zhu-Hua^{1,2,3+}, ZHANG Guo-Qing¹, YANG Jing^{1,4}, FU Chuan¹, ZHANG Guo-Qiang^{1,5}

¹(Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

²(Graduate University, The Chinese Academy of Sciences, Beijing 100190, China)

³(Key Laboratory of Knowledge Processing and Networked Manufacturing (Hu'nan University of Science and Technology), Xiangtan 411201, China)

⁴(The Research Institution of China Mobile, Beijing 100053, China)

⁵(School of Computer Science and Technology, Nanjing Normal University, Nanjing 210012, China)

+ Corresponding author: E-mail: liaozhuhua@ict.ac.cn

Liao ZH, Zhang GQ, Yang J, Fu C, Zhang GQ. Scheme of relational routing for querying semantic media on Internetwork. *Journal of Software*, 2012, 23(10): 2760-2771 (in Chinese). <http://www.jos.org.cn/1000-9825/4176.htm>

Abstract: A novel relational routing scheme is proposed to solve the problem of querying and aggregating semantic media content on the network at an effective pace. The scheme can route the semantic media query and quickly backtrack relevant content from the network based on the named media and semantic association. First, the models of semantic media and relational query are introduced, along with the necessary data structures and algorithms for relational routing protocol. In particular, the relational matching algorithm, the procedure of relational routing and the approach for avoiding incomplete response of pending relational query are elaborated on. Second, the key problems of relational routing are discussed, such as media naming, query preference and the application. Finally, the experimentation of relational routing model and algorithms on real platform are made. Results show that the scheme is highly efficient for retrieving semantically relevant content and provides an effective approach for semantic aggregation of distributed and dynamic media.

Key words: content networking; name-based routing; relational routing; semantic query; data aggregation

* 基金项目: “核高基”国家科技重大专项(2011ZX03002-002-03, 2011ZX03002-004-03); 国家自然科学基金(61174152, 61100178, 61100054); 北京市自然科学基金(4112057);

收稿时间: 2011-01-18; 定稿时间: 2011-12-31

摘要: 针对互网络中媒体语义关联内容的快速查找和聚合方面的问题,提出了一种新的面向网络的关系路由方案.该方案在命名媒体和语义关联的基础上对网络中的语义关联请求进行路由,然后快速返回关联内容.首先介绍了语义媒体模型和基于网络的关系查询模型,设计了关系路由通信协议的数据结构和算法,尤其是对关系匹配算法、路由过程以及关系请求的非完备性返回避免方法进行了重点介绍.然后,对关系路由的关键问题,如媒体命名、查询偏好以及应用等方面的问题进行了讨论和分析.最后,在实际环境中对关系路由的模型和算法进行了实验.结果表明,关系路由方法能够快速获取语义关联内容,并为分布、动态的媒体语义聚合提供了一条有效的途径.

关键词: 内容网络;基于名字的路由;关系路由;语义查询;数据聚合

中图法分类号: TP311

文献标识码: A

互网络已经成为媒体内容共享和发布的主要平台,然而随着媒体内容的数量与日剧增,在满足全球用户需求的过程中,媒体内容出现了数据碎片和位置分布等许多新的特征,使得人们获取语义完整的媒体内容以及语义关联内容非常困难;而且随着用户个性化要求的多样化,还出现了媒体数据的语义碎片化趋势,使得当媒体内容快速增长时,用户满意度不升反降.虽然领域本体^[1]、本体映射^[2]以及语义 Web^[3]等技术为媒体数据源所拥有的媒体数据提供了语义聚集、关系推导的能力,但是却难以适用分布于互网络中大量跨域的媒体内容,因为其媒体数据分布在世界各地,在系统开放的整个生命周期中,各媒体数据源自身数据及之间的语义关联经常发生变化.现有的分布式数据查询技术一般需要先对各数据源局部的数据模式进行集成,维护一个全局的逻辑数据模式.但是在大规模网络中,维护一个逻辑统一的全局模式具有很大困难,任何一个成员数据库局部模式的加入、退出和修改都需要对全局模式进行更新.另外,在大规模、复杂的语义数据模式中,个性化内容的查询效率也比较低.总之,目前对互网络中大规模动态分布的媒体内容查询的挑战主要有:

- (1) 为动态分布内容建立索引;
- (2) 复杂查询在动态网络中的转发有效性和可达性;
- (3) 实时、完备地返回查询结果.

本文首先利用本体和语义 Web 技术定义媒体数据的语义模型,以实现各媒体数据源的统一访问和查询;然后,基于统一命名和关系定义的媒体内容,提出一种灵活的关系路由和转发机制,即主要基于参考关系和包含关系实现查询的路由和转发,以实现快速查询和聚合动态分布的语义媒体内容.这是互网络中语义媒体内容获取和聚合方面急需解决的关键问题,甚至是将互网络中的数据海洋变成能够为用户提供个性化、高效服务的知识海洋的技术基础.

本文第 1 节介绍相关的研究工作,对现有分布式数据查询机制进行概括.第 2 节对网络媒体的语义模型以及关系查询进行定义.第 3 节介绍关系路由模型和相关算法.第 4 节对关系路由的一些重要问题进行讨论.第 5 节对关系路由模型进行实验,并对其效率做出评价.最后是总结和展望.

1 相关工作

针对互网络中大规模分布数据的查询,人们提出了分布式数据查询、Web 数据搜索引擎和自组织网络的数据查询机制和方法等.

分布式数据查询机制的实现方法,一般是在分布式数据查询之前对分布的局部数据模式进行集成,建立一个全局数据模式,然后在查询时根据全局数据模式与局部数据模式的映射将查询请求进行划分,重写成多个子查询,每个子查询对应一个局部的数据模式.由于媒体数据的描述和存储形式不仅存在结构化的关系数据,而且还存在半结构化的 XML 数据,因此解决分布式数据查询的方法主要有两大类:

- 一种是基于关系数据库查询理论的分布式数据查询方法,如多数据库查询语言 SchemaSQL^[4]、联邦数据库^[5]和中间件信息系统(MBIS)^[6]等;
- 另一种是针对分布式 XML 语义数据的查询方法.如在文献[7]中,作者借鉴 MBIS 的结构提出了一种分布式查询引擎 DRAQ(扩展的 ARQ,ARQ 见 <http://jena.sourceforge.net/ARQ/>)以对分布的 SPARQL^[8]资源

库进行访问和查询;在文献[9]中,作者使用了特定的索引结构实现对分布的 SeRQL^[10]资源库的查询。

但无论数据是结构的还是半结构的,在大规模网络中,其查询都需要在查询之前建立全局的数据模式.然而,要在大规模网络中维护一个全局的数据模式是相当困难的,尤其对于跨域、动态变化的媒体内容,建立或维护全局的数据模式则更加困难.

对于 Web 网络中分布数据的查询以及关联关系的搜索等应用,则一般需要先基于爬虫等技术采集各 Web 网站的数据,然后对其进行数据分析,以建立对象库、索引库和关系库等.如提供人物关系搜索服务的人立方关系搜索引擎(<http://renlifang.msra.cn>),从超过 10 亿的中文网页中自动地抽取出人名、地名、机构名以及中文短语,并通过算法计算了它们之间可能存在的关系.

自组织的内容网络可以看成是一个动态的大规模分布式数据库,而如何在这种网络中进行数据查询,是一个颇具挑战的问题.人们在 P2P 网络技术研究中提出了基于路由机制的数据查询方法,如非结构化 P2P 网络中有两种方式:一种是对数据查询消息的洪泛^[11]、随机游走^[12]和超级节点路由^[13]方法.洪泛或随机游走等方法的查询开销比较大,查询结果的完备性无法保证;超级节点路由方法是将少数能力强的节点作为超节点,它们负责查询消息的路由,而普通节点则将自己共享的数据信息发布到与其连接的超节点上,不参与查询消息的转发.由超节点组成的网络主要采用洪泛方法,因此与纯洪泛或随机游走方法相比,数据查询的效率得到了提高,但大规模的查询以及数据的动态更新容易使超节点产生性能瓶颈和单点故障;另一种是源内容语义的广播和查询分解的方式^[14],由于要广播所有语义信息以及查询分解的计算复杂度较高,从而影响了查询性能.对于结构化 P2P 网络,常用的是基于分布式散列表(DHT)对数据进行定位,如 Chord^[15],Pastry^[16]等;另外一种是基于语义构建和维护层次语义 Overlay 网络(SON)^[17],即,对内容语义相似的节点进行聚合,查询时对用户请求进行分类转发.总之,结构化 P2P 网络的查询路由方法虽然效率比较高,但需要建立严格的网络拓扑结构并且维护代价高,不适合节点或数据变化比较频繁的动态网络环境.目前,在传感器网络中,动态分布数据的查询仍然采用洪泛和随机游走等方法或对其进行局部改进^[18-20],因此查询效率和开销等问题依然难以解决.

近些年来,为了获取动态分布的内容,人们提出了基于名字的路由方法(本文简称为名字路由机制).其主要原理是:当数据源在发布新的内容时向路由节点进行内容通告,用户则根据所需内容的名字发送请求到离用户最近的路由节点;然后,路由节点根据名字进行匹配和转发,直到从路由缓存中或数据源中获取名字对应的内容并原路返回.人们对名字路由机制已进行了长期研究,提出了 TRIAD^[21],DONA(data-oriented network architecture)^[22]和 CCN(content-centric networking)^[23]等基于名字路由机制的全新网络架构,它们对其内容命名、节点模型、路由协议和内容匹配算法等进行了研究,实现了通过名字直接在网络中获取内容的有效方法,解决了网络中内容获取的准确度和时效性、数据传输的有效性、数据位置的依赖性问题.但是,目前名字路由机制只实现了“一对一”的数据获取,即一个请求只获取一个内容,还不支持复杂、完备的语义关联内容查询.但 CCN 这种层次化命名机制和基于名字的组播路由方式为处理互连网络中语义媒体查询提供了一种新的研究方向.

本文基于 CCN 这种简单的层次化命名,研究一种全新的语义媒体查询方法,以解决互连网络中动态分布内容的复杂语义查询以及查询结果完备返回的问题.

2 网络媒体的语义模型

2.1 媒体语义关系定义

网络中各媒体数据源具有自治管理、安全控制等优点,但要实现网络中各自治数据源的统一访问和查询,则首先需要对数据源自身数据进行统一命名和关联关系的统一定义.目前,本体技术为数据和关联关系的统一命名和定义提供了有效途径,其存储可以采用数据库模式或者 RDF 与 OWL 语言描述的文件形式,从而形成能够被统一访问的媒体数据库.本体可以表示为一个二元组 $(ID, Axiom)$ ^[2],其中, ID 是定义媒体涉及的词汇集合,且 $ID = C \cup R \cup I$,其中, C 表示媒体的概念集合, R 表示媒体之间的关系集合, I 表示媒体实例集合; $Axiom$ 是指媒体之间语义关系推导公理的集合.因此,对于各媒体数据源中媒体对象及对象之间的关系,可以采用本体技术进行定

义.为了便于灵活地查询媒体内容,媒体对象的名字采用概念的层次结构命名.即:一个媒体对象的命名包括所属类名和数据名字,这样有助于在网络中实现基于前缀或主题的查询等;而媒体对象之间关联关系的定义则在领域本体已定义的关系的基础上,根据媒体对象特有的关联特征增加常用的关系类型.

定义 1(后继(successor)). 表示为 $Suc(m_i, m_j)$,即声明媒体 m_j 是媒体 m_i 的后继;对于视频,则表示为 m_i 播放完之后可以立即播放媒体 m_j .另外,也可以称 m_i 是 m_j 的前驱(precursor),表示为 $Pre(m_j, m_i)$.

定义 2(组合(composition)). 表示为 $Comp(m_{i1}, m_{i2}, \dots, m_{in})=m_i$,即,声明媒体片段 $m_{i1}, m_{i2}, \dots, m_{in}$ 的组合在语义上等价为 m_i ,并且 $m_{ik}(1 \leq k \leq n)$ 为 m_i 的组成部分.

媒体的组合可以分为时间组合、空间组合和混合组合这 3 种:

- 时间组合表示为 $Comp_t(m_{i1}, m_{i2}, \dots, m_{in})$,其中, $m_{i1}, m_{i2}, \dots, m_{in}$ 之间是前驱关系,即, m_{i1} 是 m_{i2} 的前驱, m_{i2} 是 m_{i3} 的前驱, $m_{i(n-1)}$ 是 m_{in} 的前驱;
- 空间组合表示为 $Comp_s(m_{i1}, m_{i2}, \dots, m_{in})$,即, $m_{i1}, m_{i2}, \dots, m_{in}$ 之间同时呈现或播放,但在空间布局上的位置不同;
- 根据时间和空间的混合组合又分为时空混合组合和空时混合组合,其中,
 - ◊ 时空混合组合表示为 $Comp_{ts}(comp_t(m_{1a}, m_{1b}), comp_s(m_{2a}, m_{2b}), \dots)$;
 - ◊ 空时混合组合表示为 $Comp_{st}(comp_s(m_{1a}, m_{1b}), comp_t(m_{2a}, m_{2b}), \dots)$.

定义 3(补充(supplement)). 表示为 $Sup(m_i, m_j)$,即,声明媒体 m_j 是媒体 m_i 的补充内容或附加内容.

定义 4(替代(substitute)). 表示为 $Sub(m_i, m_j)$,即,声明媒体 m_j 可以替代媒体 m_i .

上述补充和替代关系具有非对称和传递性特征,即:

- $Sub(m_i, m_j)$ 成立并不等价于 $Sub(m_j, m_i)$ 成立;
- 但如果 $Sub(m_i, m_j)$ 和 $Sub(m_j, m_k)$ 成立,则有 $Sub(m_i, m_k)$ 成立.

定义 5(副本(duplicate)). 表示为 $Dup(m_i, m_j)$,即,声明媒体 m_j 是媒体 m_i 的副本,或者说媒体 m_i 和媒体 m_j 在语义和媒体格式上等价.副本关系具有对称和传递特征.

分布在网络中的数据源除了具有大量的媒体内容以外,还具有自身媒体内容的关联关系定义.对于跨域的媒体对象组成的语义关联图,则是一个由节点(表示媒体类名或实例名)和边(表示关系)组成的混合图,如图 1 所示.

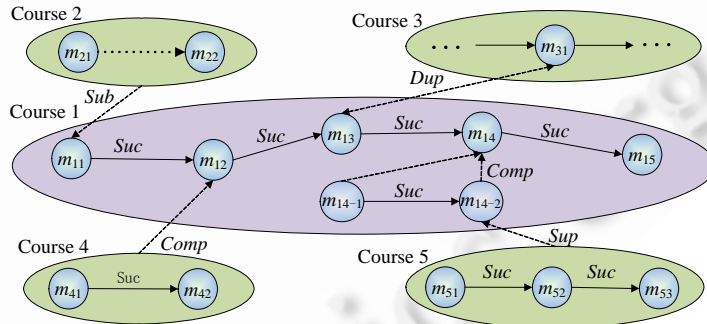


Fig.1 Semantic association graph of media

图 1 媒体语义关联图

2.2 关系查询模型

关系数据库和 XML 数据都可以建模成数据图,图中的节点分别对应关系数据库中的元组(由各属性列组成)和 XML 数据的内容(由 XML 数据中包含的 text 以及属性值组成),图中的边分别对应关系数据库中元组之间的主、外键关联和 XML 数据中的结构信息(由 XML 数据标签间的嵌套关系组成)^[24,25].数据查询的本质是根据数据的属性约束或者关联约束进行查询.面向网络的最有效的查询过程是:先将用户的关系查询请求路由到网络中可能的节点,然后再在可能的节点中进行关系查询和匹配,返回语义关联的内容.

基于网络的数据查询模型需要考虑请求的复杂性和传输、路由的方便性,另外就是路由节点的缓存和数据匹配.网络中数据查询的结果具有如下特征:

- (1) 一对一,即只返回任一节点上的一个对应数据即可.因此,可以在可能的节点中对查询命令进行任播,然后等待查询结果返回.如果对查询响应时间有较高要求,则可以在可能的节点中对查询命令进行组播,然后接收最先返回的结果.如,基于名字的路由机制采用的就是这种方式;
- (2) 一对多,即返回任一节点上的多个匹配数据或者多个节点上的所有匹配数据.对于前者,也可以采用任播方式;对于后者,则一般采用组播的方式返回多个节点的匹配结果.但在路由中,这种情况会在查询请求的定义、转发、匹配以及返回结果的缓存、重复性过滤、完备性等方面引入新的问题.

为方便语义查询请求的传输和路由,并能够在路由节点进行匹配,本节定义一种基于参考关系的查询模型,除了指定语义约束以外,还指定与结果有关的路由导向信息.根据语义约束,关系查询模型分为两类,即关联约束查询模型和属性约束查询模型.

- (1) 关联约束查询模型:关联约束查询模型是一个三元组 (RO, R, RF) , RO 是参考对象名字, R 是关系名字, RF 是参考对象与期望获取的结果之间的关联逻辑运算表达式.在路由中,统一用符号串 $Reference_r$ ($Relation_1 OP_L Relation_2 OP_L \dots$)表示,其中, $Reference$ 是参考对象的层次命名, $Relation_i$ 是关系名字, OP_L 是逻辑运算符(如 AND, OR, NOT 等).如果查询媒体 n_i 的后继或者媒体 n_i 的组合和补充媒体,则分别表示为 $n_i_r(Suc)$, $n_i_r(Comp OR Sup)$;
- (2) 属性约束查询模型:为了限定内容分类层次中的查询范围需指定所属类名,查询时只在所指定的类别中进行查询,因此,其查询模型是一个四元组,即 (CO, PF, F, V) , CO 是查询结果所属类名, PF 是分类的属性之间的关系运算表达式, F 是指定的属性名集合, V 是属性名对应的值的集合.在路由中,统一用符号串 $Class_f(Feature_1 OP_M V_1 OP_L Feature_2 OP_M V_2 \dots)$ 表示,其中, $Class$ 为分类名字, $Feature_i$ 为属性名字, OP_M 为关系运算符(如 =, >, < 等), OP_L 为逻辑运算符, V_i 为属性的限定值;
- (3) 返回结果:用户的查询请求无论是关联约束查询还是属性约束查询,其返回结果都是一对多的,因此,返回结果的形式表示为 $\{\{name_1, name_2, \dots\}_{site1}, \{name_1, name_2, \dots\}_{site2}, \dots\}$.

3 关系路由模型

3.1 关系路由数据包

现有 IP 网络体系结构主要以“协议”为中心,具有沙漏型的结构,以维持路由协议在网络中路由的简单性、鲁棒性和可扩展性.同样,我们对于关系路由仍以“协议”为中心,将关系查询请求和返回数据分别映射到路由协议中的关联兴趣包和关联数据包.为了兼顾基于名字的路由,即在获取关联内容后使用名字路由机制获取名字对应的媒体内容,我们参考 CCN^[23]中有关兴趣包和数据包的格式分别定义关联兴趣包和关联数据包.关联兴趣包主要封装用户发送的关系查询请求,但关联兴趣包中具体字段的数据格式具有不同的定义,其格式见表 1(a).关联兴趣包名字(name)的具体定义为第 2.2 节中的关系查询请求格式.其中,Selector 字段可以携带用户的偏好约束,如限定关系名字的属性等;Scope 字段可以让用户限定返回关联内容的数量、查询超时时间等;Detector 字段是避免关联兴趣包转发循环.关联数据包的返回内容主要是与关系查询请求匹配的媒体语义关联对象的名字列表,而不是媒体内容对应的数据,具体表项的定义见表 1(b).

Table 1 Definition of data packets

表 1 数据包定义

(a) 关联兴趣包		(b) 关联数据包	
Name	名字_r(关系查询约束)	Name	名字_r(关系查询约束)
Selector	偏好约束,发布者过滤	Signature	名字对应内容的发布者
Scope	返回关联数量、超时时间	Signed info	签名信息
Detector	Nonce (避免转发循环)	Data	关联内容信息列表

关联数据包的名字与对应的关联兴趣包的名字相同,Signature 字段是登记“参考对象或分类”的发布者,而 Signed info 字段携带保证数据安全的签名信息。

3.2 关系路由节点模型

关系路由节点需要考虑返回数据的缓存、重复数据的过滤、返回结果的完备性和请求的匹配等问题。关系路由转发节点除采用了与 CCN 类似的基本数据结构外,如 ContentStore(数据缓存)、PIT(pending interest table)表和 FIB(forward information base)表,还增加了 RRT(response relation table)表以便结果的完备返回和冗余抑制。ContentStore 除了缓存数据名字对应的内容,还缓存返回的关联内容。缓存中的关联内容主要通过“关联数据包的名字与发布者(或网站网址)”来唯一标识,缓存内容的替换策略则采用通用的 LRU 算法。FIB 表主要存储数据源发布的媒体内容的名字(采用层次命名)和名字所对应的 Facelist(入口列表),而不存储名字之间的关联信息,因而所占存储空间与 CCN 相同;但在关系路由中,需要完成复杂的语义查询和结果返回。

3.2.1 未决查询的处理和完备性返回

由于关系路由中结果返回具有“一对多”的特征,为了保证接收节点能够接收并返回不同关联内容列表,未决查询的处理是关系路由的关键问题。PIT 表暂存等待结果的关系查询请求,即关联兴趣包的名字。PIT 表的格式如图 2 所示,在设计中考虑如下关键问题:

- (1) 表项删除时机。当有关联内容返回时,并不马上删除 PIT 对应的关系查询请求表项(若马上删除,则对于每个请求最多只能返回一个数据包),而是设置一个删除时机选项,方案如:对等待结果的关联兴趣包设置最大超时或者设定允许返回关联内容列表的最大数目;当超过一定时间或者返回内容项超过最大数目时,自动删除 PIT 对应表项,并向其他路由节点通告,从而保证返回多个不同数据源的相关内容列表。返回的每个关联数据包只包含一个数据源中全部或部分关联内容名字列表;
- (2) 非完备性返回的避免机制。当不同用户发出相同关系查询请求时(即关联兴趣包的名字相同),可能会出现部分后请求的用户无法收到或只能返回一部分关联内容的情况。如,用户 A 发出 $n_{i,r}(Suc)$ 的查询请求之后,在关联内容已经返回一部分或全部返回,但 PIT 表中对应关联兴趣包的表项还没有删除之前,用户 B 发出同样的查询请求后,PIT 表会将相同查询请求合并处理。因此,可能会在用户 B 发出请求之前,已经返回一部分或者全部数据包,用户 B 只能接收后一部分返回数据包或者没有接收而空等。我们把这种情况称为非完备性返回。为避免这种情况的发生,我们在路由转发节点中增加 RRT(response relation table)表。RRT 表在节点转发查询请求后对返回数据包进行记录,主要存储返回关联数据包的头信息,包括关系查询请求名字和发布者(publisher)信息(假定媒体发布站点与发布者一一对应,用以确定关联数据的来源)等,并且 PIT 表中每个表项关联多个 RRT 表项,RRT 的格式如图 2 所示。路由节点对 RRT 表的具体操作算法如下:
 - (a) 当部分关联数据包已经返回后,若有新的用户发出相同关系查询请求,则在 RRT 表中记录其入口(face)和到达时间(arrival time);当返回剩余关联数据包时,记下关联数据包信息中包含的 Publisher;
 - (b) 对于已经登记但未返回的关联数据包,首先查询缓存,如果缓存中存在对应的关联数据包,则返回,并在 RRT 中对应入口项记下返回关联数据包的名字和 Publisher 信息;如果缓存中不存在,则查看同一查询请求不同入口的返回情况,如果存在有差异的非完备性返回,则重新转发请求;
 - (c) 检查已经超时的 PIT 表项,对于已经返回所有在超时之前接收到的关联数据包的关系查询请求,删除其对应表项中的入口以及该入口相对应的 RRT 表项。如果对应的 PIT 表项的入口列表为空,则删除对应的 PIT 表项;如果非空,说明存在非完备性返回,则重新转发请求。

引入超时机制虽然满足了关联兴趣包和关联数据包的一对多属性,但延长 PIT 表项的删除时间会影响后续相同请求的有效响应。因此,我们在上述算法步骤(b)中通过对非完备性返回进行差异性检测,克服了超时机制带来的这种影响。

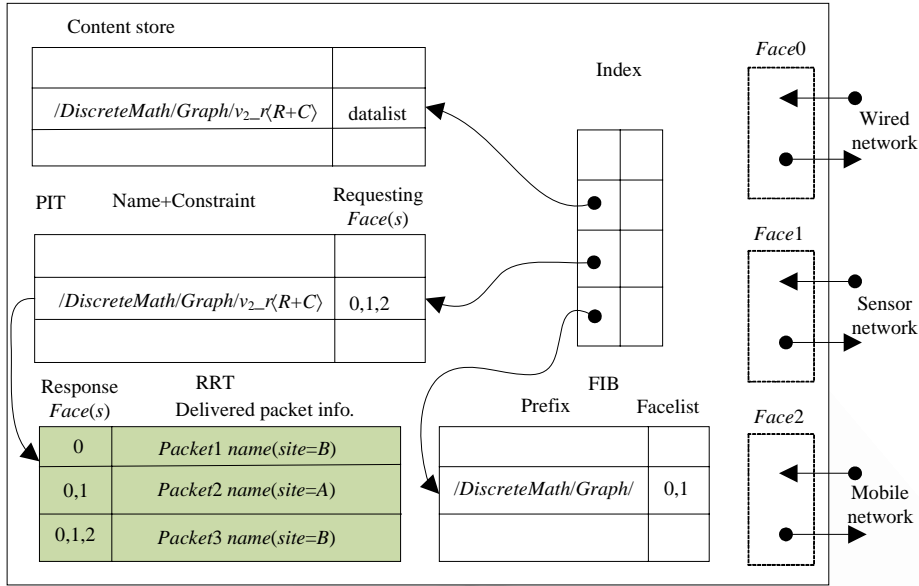


Fig.2 Data structure of forward node for relational routing

图2 关系路由转发节点数据结构

3.2.2 关系路由过程和匹配算法

当用户要获取特定对象或特定类别的关联内容时,关系查询请求路由的主要过程是:首先向邻近路由节点发送关联兴趣包;然后通过路由节点将其转发到潜在的路由节点,并在路由节点的缓存中进行匹配,如果匹配结果不能满足完备性返回,则再转发关联兴趣包,直到各潜在的数据源节点;当各数据源节点接收到关联兴趣包后,经过语义查询后将用户所需的关联内容名字列表封装进关联数据包,然后原路返回;各路由节点再根据 RRT 对应表项的记录决定是否返回,如果返回,则根据 PIT 对应表项的入口决定返回的下一节点,然后在 ContentStore 中缓存关联数据包,直到缓存内容被替换掉.图 3 显示了用户关系查询请求总体的路由和匹配过程.

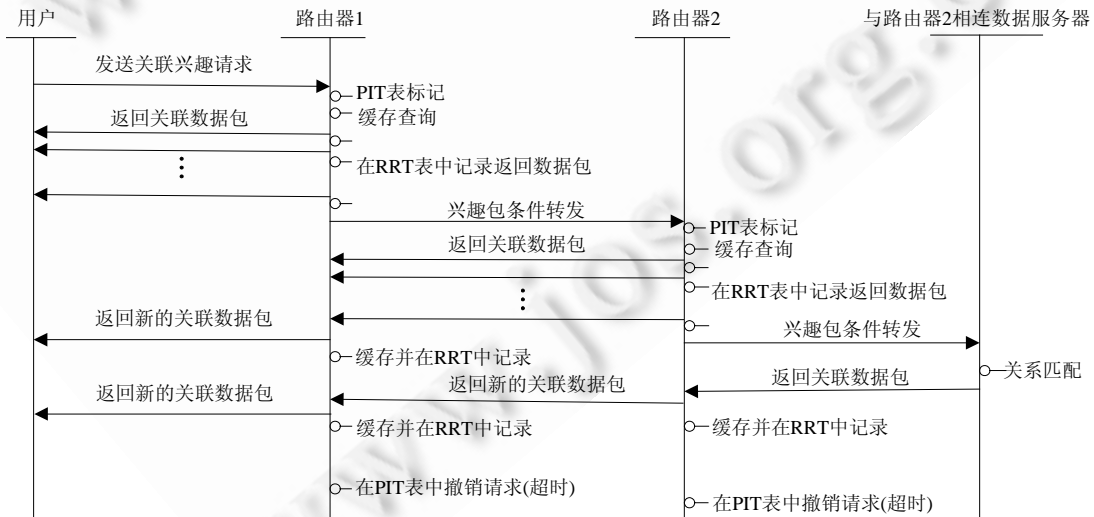


Fig.3 Routing and match process of association interest packet

图3 关联兴趣包的路由和匹配过程

路由器节点对关系查询请求的匹配主要采用最长前缀匹配策略,关联兴趣包发送后,路由节点中关系匹配

算法如下所示:

- (1) 接收节点先在 PIT 表中对关联兴趣包的名字进行匹配,如果匹配成功,则在 PIT 相应表项中增加相应入口.查询 RRT 表确定是否有对应表项,如果有对应表项,则根据第 3.2.1 节中的操作算法查询缓存、登记并返回数据包;如果 PIT 表中没有查询请求对应的表项,则在 PIT 表中增加相应表项和入口;
- (2) 在 FIB 表中,根据关联兴趣包名字中提供的参考对象名或类别名进行匹配(因为对象或类别的关联信息一般在同一位置),如果匹配成功,则根据对应表项的 Facelist 字段列出的出口对关联兴趣包进行转发;否则,把关联兴趣包丢弃;
- (3) 当有数据包返回时,首先在 RRT 表中查询对应请求表项,对没有登记返回的请求入口进行回传,并在 RRT 表中进行登记,然后在 ContentStore 中缓存.

源节点的关系匹配过程是:当最终数据源节点收到关联兴趣包后,它根据参考对象名或类别名以及约束条件查找和匹配已知对象或类别所关联的数据集合,将所有关联对象名字列表作为关联数据包的数据内容,关联兴趣包的名字作为数据包的名字项,并与发布者信息等封装,原路返回关联数据包.

在路由节点中,首先在 RRT 表中匹配返回数据包,而不是先匹配 ContentStore 里的内容.这是由于存在“一对多”的关联关系特征,用户和中间路由节点事先都不知道网络中副本的分布和数目,也不确定 ContentStore 中缓存的内容是否完备,因此需要先在 RRT 表中匹配并查询缓存中的关联内容并返回和登记,即,RRT 相当于一个列表用来存储返回的内容.另外,用户对单个数据的获取则可以只用数据名字作为兴趣包的名字,数据匹配中只需使用 PIT 表、FIB 表和 ContentStore 数据结构并采用名字路由匹配算法即可完成.关联数据包的返回要保证一个关系查询请求从每个对应数据源至多只接收一份.本文主要通过发布者签名信息进行抑制,当一个路由节点同时接收到多个名字与发布者一致的关联数据包时,则最多只向每个入口返回一份.具体如图 4 所示,用户查询 n_3 的后继,即发送 $n_3_r(Suc)$ 关系查询请求,两个数据服务器都存在关联内容,而且路由节点 F 也缓存相关关联内容,因此,用户到可能的数据源有 3 条路径($u_1-E-C-A, u_1-E-C-B, u_1-E-F$).这样,路由节点 E 会先后收到 3 份返回结果($\{n_{10}, n_{31}\}, \{n_4\}, \{n_4\}$),其中,两份 $\{n_4\}$ 结果是相同的,因此,路由节点 E 必须过滤重复的返回内容.

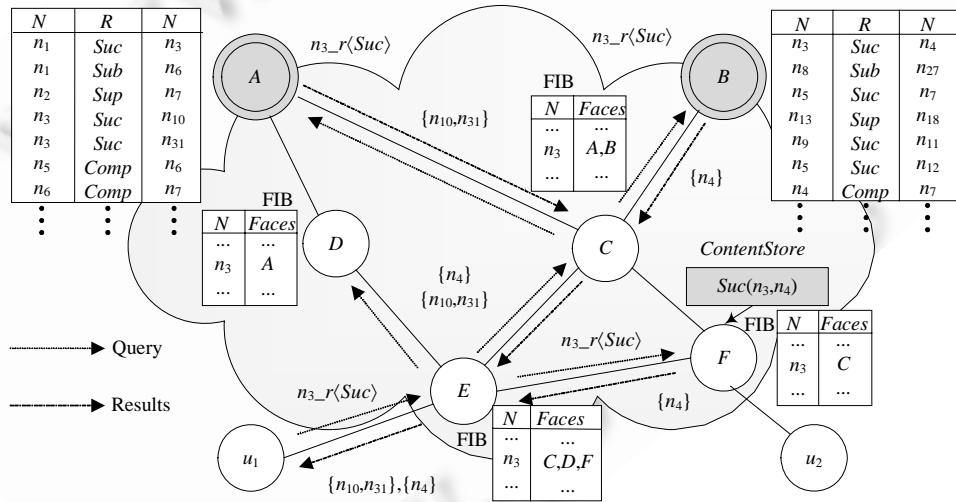


Fig.4 An example of relational query

图 4 关系查询实例

4 讨论

4.1 关系查询请求的命名问题

在关系路由模型中,关系查询请求的命名对具有“一对多”特征的返回结果的正确性和完备性具有重大影

响.通常,在大规模网络中要获取完备的关联信息,关系查询请求包的名字标识需要与位置无关,并且名字唯一标识数据.目前,位置无关且保证唯一的数据命名方法已有很多种,文献[26]对位置无关的数据命名方法进行了详细介绍,如 PURL(持久性统一资源定位器)、URC(统一资源引用符)以及 DONA 采用的扁平命名方式^[22]等.然而对于大规模互连网络中海量的媒体数据,人们一般会采用分类和分层的方式对其进行存放,建立了不同媒体类别之间的语义关联关系,因此需要关联请求名字命名具有相对的灵活性.CCN 采用了层次命名结构,用户需要提供的数据请求名字包括两部分:全局路由名字(globally-routable name)和组织名字(organizational name),其中,全局路由名字指定内容所属域名,组织名字包括组织层次名字和具体数据的名字,而 DONA 只定义可唯一标识的具体数据名字.这两种命名方式都不适合关系查询请求的命名,因此,本文的关系查询请求命名中只给定类别名字和数据名字的组合的前缀部分.即,可以不给定具体数据的名字,数据源端关系匹配时,匹配已知媒体类名对应关联关系的内容,其结果可以是具体数据名字的集合,也可以是类名.因此,该命名方式可以依据这种层次化的名字实现跨域转发,使得查询请求到达所有可能的数据源.其名字的匹配方法仍可以采用最大前缀匹配方法,只是当名字限定的范围越小,则查询范围越小,返回的数据包数量也会越少,但速度会加快.

4.2 偏好约束问题

关联兴趣包中的 Selector 字段可以携带用户的偏好约束等,但偏好约束会对关系匹配结果产生影响.即,关联兴趣包名字相同但偏好不同,则返回的一对多的关联内容可能不相同.因为相同名字的请求在网络中增加偏好所对应的内容集,是不增加偏好所对应的内容集的子集.直至今日,基于名字的路由机制也还未考虑偏好约束问题的解决方法.名字和关系相同但偏好不同,所返回的关联内容集合可能出现包含、交叉、不相交或无法判断等情况,因此,我们的解决方法是:路由节点首先分析偏好条件之间的语义关系,如果新约束条件与已有约束条件是被包含的关系,则缓存内容必包含用户所需内容,可进一步根据新的偏好进行过滤返回;对于交叉、不相交和无法判断的情况,则认为不同的请求,重新对新的请求进行转发.另外,在处理偏好约束时,偏好约束越复杂,则关系路由的数据结构和匹配算法越复杂.因此,在关系路由协议设计中应该兼顾功能和性能的折衷.

4.3 关系路由的应用

本文提出的关系路由机制以及 CCN 等提出的名字路由机制相对于现有网络的 IP 路由机制(我们称其为地址路由机制)是一种全新的路由方式.从对数据路由的角度来看,它们的区别主要是:关系路由机制(包括名字路由机制)的复杂性一般只与内容数量有关,而与网络节点数基本无关.因此,网络中内容数量须有限,节点数可以无限大;地址路由机制则刚好相反,地址须有限,但内容数量可以无限大.因此,关系路由或名字路由比较适用于内容变化较大、节点数远大于内容数量的网络环境,即,用于内容比较稀疏、动态变化的网络环境更有优势.如,对于由很多传感设备互联的大规模物联网环境,其生产总的的数据量(或者在一段时间之内)并不多,即大部分节点并不总是产生数据而主要是作为数据的转发节点.对于这种环境,如果采用传统的关系数据查询方法,则首先需要对每个节点进行查询遍历,然后构造全局的数据模式,但对于物联网环境数据动态变化比较大,维护全局数据模式比较困难;而采用关系路由方法则可以对节点新产生的数据进行动态发布,用户获取关联内容也无需全局遍历,因此可以比较快速地查询并获取相关内容.

相比于名字路由机制 CCN 仅用于单个内容的获取,关系路由机制则主要适用于一对多(多个节点、多个数据)的动态分布数据的语义查询,而且在路由节点上只传输和占用空间小的关联内容而不是数据的二进制内容.另外,关系路由机制充分利用了组播方式转发查询和聚合结果,因此在开放的互连网络中,对于动态分布内容的语义查询和聚合方面具备较强的应用潜力.

5 实验

为了验证关系路由方法的可行性及其性能,我们采用 Java 开发语言和 MySQL 数据库实现了关系路由的原型系统,并进行了性能测试.关联兴趣包和关联数据包的转发主要采用 TCP 协议传输.由于对半结构化的数据(如 RDF 等描述的数据)可以通过转换部件将其查询结果转换为结构化数据后返回,因此在实验中,服务器上媒

体数据的语义关联采用结构化数据库存储.然后,我们在局域网络中通过 68 台 PC 机组成的集群上构建了由 1 020 个路由节点(用进程模拟)互联的 Overlay 网络拓扑,其中,路由节点随机连接 1 054 个多媒体课程数据源,PC 机的硬件配置是 Intel Pentium 4 CPU 和 512M 到 1G 的内存组成.实验中,用户节点可以选择与任一中间路由节点相连,用户与服务器的最大平均跳数可以达到 6.在每一个数据源中,存储了 30 门课程约 210 个媒体对象中的部分或所有语义关系数据.所有课程之间、媒体对象之间建立大约 2 000 个关联关系,主要包括多媒体对象(主要是指每个多媒体视频分段)、分类以及对象、类别之间的关联关系.多个视频分段按照课程知识点之间的关系,如补充(*Sup*)、替代(*Sub*)等关联关系组成,课程之间建立学习顺序等逻辑关系,如前驱(*Pre*)、后继(*Suc*)等.另外,课程、章、节对应的媒体对象建立组合(*Comp*)关系.在不同服务器上,相同视频分段认为是副本(*Dup*)关系,对于对象名字和关系的标识采用统一的命名和关系定义.

图 5 显示了用户与服务器的平均跳数为 6 时(仅数据转发情况下平均往返时延 RTT 为 531ms),媒体副本数(相同媒体分布在不同服务器的数量)分别与第 1 个数据包返回及关联数据完备返回的总路由时间的关系.在这组实验中,当副本数从 2 增加到 16 时,平均总路由时间从 2.2s 到 8.6s 增长,而第 1 个数据包的总路由时间基本不变.从图 5 中可以看出,副本数越多,总路由时间越长,但并非呈线性关系,而是随着副本数的增加,时间增长速度会越来越小.这是因为数据源比较分布,副本数增加时用户节点到各副本所在节点的路径分支数也在增加,请求转发及各副本所在数据源的关联查询的并发性都得到提高.

在实际互网络中,一般平均路由直径(平均跳数)比较小^[27],而如果关系路由建立在应用层之上,则平均路由跳数更小.设整个应用网络共有 N 个节点,并且节点跳数最大度为 d ,根据文献[28]中的计算公式,则平均路由跳数为 $(\log_d N - o(1))$.即,对于最大度为 7、节点规模达 576 万余个节点的网络,其最大平均路由跳数约为 8.图 6 显示了平均副本数为 3 时,关系查询请求转发跳数分别与第 1 个数据包返回及结果完备返回时的总路由时间的关系.从实验中可以看出,当跳数从 1 增加到 8 时,第 1 个数据包返回及结果完备返回的各自总路由时间都会同步增长,其中,结果完备返回的平均总路由时间从 0.6s 逐步增长到 3.55s.但从图 6 中可以看出,随着跳数的成倍增加,总路由时间则以较小速率增加.这主要是因为总路由时间除了包括路由节点的查询时间外,还包括数据源节点上的查询时间,本实验中各数据源中的查询时间平均为 146ms.

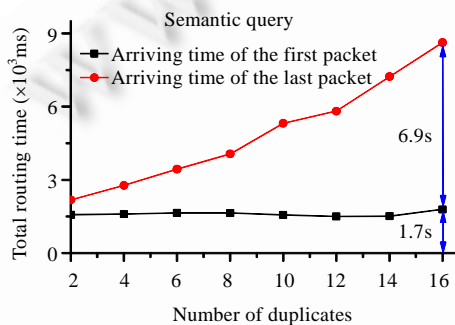


Fig.5 Relationship between duplicate factor and total routing time

图 5 副本数与总路由时间的关系

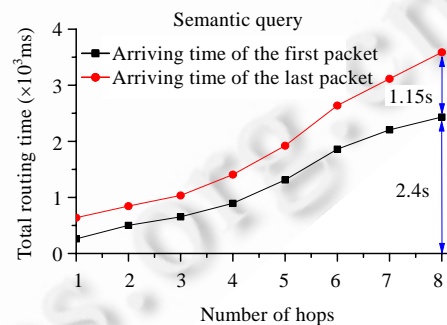


Fig.6 Relationship between forward hops and total routing time

图 6 转发跳数与总路由时间的关系

6 总结和展望

本文在媒体统一命名和关系定义的基础上,结合语义 Web 和本体技术,提出了一种适用于互网络中语义媒体查询的新型关系路由方法,并设计和实现了关系路由协议和具体算法,为大规模网络中媒体内容的语义查询提供了快速、有效的途径,并支持跨域、动态和分布的媒体数据的语义查询和内容聚合.另外,本文提出的关系路由机制还兼容基于名字的路由方式.在未来的工作中,我们将在关系路由机制的应用中进一步研究内容规模与查询效率之间的关系,并进行量化和测试,以更好地指导其应用并提出性能优化方法.另外,还将对关系路

由机制进一步加以改进,尤其是其语义扩展、用户偏好约束的研究等,使其能够更好地支持大规模、动态、分布的媒体内容的个性化语义聚集、序列生成以及查询上下文自适应等功能和应用,以更有效地解决互网络中媒体语义关联内容的快速查找和聚合方面的问题。

致谢 感谢研究组边建功等同学参与了部分协议设计的讨论和实现,以及匿名审稿专家对本文提出的宝贵建议;感谢湖南科技大学知识处理和网络化制造湖南省高校重点实验室为本文提供实验环境。

References:

- [1] Navigli R, Velardi P. Learning domain ontologies from document warehouses and dedicated Web sites. *Computational Linguistics*, 2004,30(2):151–179. [doi:10.1162/089120104323093276]
- [2] Qu YZ, Hu W, Zheng DD, Zhong XY. Mapping between relational database schemas and ontologies: The state of the art. *Journal of Computer Research and Development*, 2008,45(2):300–309 (in Chinese with English abstract).
- [3] Maedche A, Staab S. Ontology learning for the semantic Web. *IEEE Intelligent Systems*, 2001,16(2):72–79. [doi: 10.1109/5254.920602]
- [4] Lakshmanan L, Sadri F, Subramanian I. SchemaSQL—A language for interoperability in relational multi-database systems. In: Vijayaraman TM, Buchmann AP, Mohan C, Sarda NL, eds. *Proc. of the 22th Int'l Conf. on Very Large Data Bases (VLDB)*. San Francisco: Morgan Kaufmann Publishers, Inc., 1996. 239–250.
- [5] Sheth A, Larson J. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Computing Surveys*, 1990,22(3):183–236. [doi: 10.1145/96602.96604]
- [6] Wiederhold G. Mediators in the architecture of future information systems. *Computer*, 1992,25(3):38–49. [doi: 10.1109/2.121508]
- [7] Bastian Q, Ulf L. Querying distributed RDF data sources with SPARQL. In: Bechhofer S, Hauswirth M, Hoffmann J, Koubarakis M, eds. *Proc. of the 5th European Semantic Web Conf. LNCS 5021*, Berlin, Heidelberg: Springer-Verlag, 2008. 524–538. [doi: 10.1007/978-3-540-68234-9_39]
- [8] Eric P, Andy S. SPARQL query language for RDF. Technical Report, W3C Working Group, 2008. <http://www.w3.org/TR/rdf-sparql-query/>
- [9] Stuckenschmidt H, Vdovjak R, Houben G, Broekstra J. Index structures and algorithms for querying distributed RDF repositories. In: *Proc. of the 13th Int'l Conf. on World Wide Web*. New York: ACM, 2004. 631–639. [doi: 10.1145/988672.988758]
- [10] Broekstra J, Kampman A. An RDF query and transformation language. In: *Proc. of the Semantic Web and Peer-to-Peer*. Berlin, Heidelberg: Springer-Verlag, 2006. 23–39. [doi: 10.1007/3-540-28347-1_2]
- [11] Ripeanu M. Peer-to-Peer architecture case study: Gnutella network. In: *Proc. of the 1st Int'l Conf. on Peer-to-Peer Computing*. Linköping, 2001. 99–100. [doi: 10.1109/P2P.2001.990433]
- [12] Gkantsidis C, Mihail M, Saberi A. Random walks in peer-to-peer networks. In: *Proc. of the 23rd Annual Joint Conf. of the IEEE Computer and Communications Societies (INFOCOM)*. Hong Kong, 2004. 7–11. [doi: 10.1109/INFOCOM.2004.1354487]
- [13] Wolfgang N, Martin W, Wolf S, Christoph S, Mario S, Ingo B, Alexander L. Super-Peer-Based routing strategies for RDF-based peer-to-peer networks. In: *Proc. of the 12th Int'l Conf. on World Wide Web*. New York: ACM, 2003. 752. [doi: 10.1016/j.websem.2003.11.004]
- [14] George K, Vassilis C. Semantic query routing and processing in P2P database systems: The ICS-FORTH SQPeer middleware. In: Lindner W, Mesiti M, Türker C, Tzitzikas Y, Vakali A, eds. *Proc. of the Current Trends in Database Technology. LNCS 3268*, 2005. 433–436. [doi: 10.1007/978-3-540-30192-9_48]
- [15] Stoica I, Morris R, Karger D, Kaashoek M, Balakrishnan H. Chord: A scalable peer-to-peer lookup service for internet applications. In: *Proc. of the SIGCOMM*. New York: ACM, 2001. 149–160. [doi: 10.1145/964723.383071]
- [16] Rowstron A, Druschel P. Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems. In: Guerraoui R, ed. *Proc. of the Middleware. LNCS 2218*, London: Springer-Verlag, 2001. 329–350. [doi: 10.1007/3-540-45518-3_18]
- [17] Arturo C, Hector GM. Semantic overlay networks for P2P systems. In: Moro G, Bergamaschi S, Aberer K, eds. *Proc. of the Agents and Peer-to-Peer Computing. LNCS 3601*, Berlin, Heidelberg: Springer-Verlag, 2005. 1–13. [doi: 10.1007/11574781_1]
- [18] Intanagonwiwat C, Govindan R, Estrin D. Directed diffusion: A scalable and robust communication paradigm for sensor networks. In: *Proc. of the ACM/IEEE MobiCom on Mobile Computing and Networking*. New York: ACM, 2000. 56–67. [doi: 10.1145/345910.345920]

- [19] Yao Y, Gehrke J. The cougar approach to in-network query processing in sensor networks. *ACM SIGMOD Record*, 2002,31(3): 9–18. [doi: 10.1145/601858.601861]
- [20] Servetto S, Barrenechea G. Constrained random walks on random graphs: Routing algorithms for large scale wireless sensor networks. In: *Proc. of the 1st ACM Int'l Workshop on Wireless Sensor Networks and Applications*. New York: ACM, 2002. 12–21. [doi: 10.1145/570738.570741]
- [21] Gritter M, Cheriton D. TRIAD: A new next-generation Internet architecture. 2000. <http://www-dsg.stanford.edu/triad>
- [22] Koponen T, Chawla M, Chun BG, Ermolinskiy A, Kim KH, Shenker S. A data-oriented (and beyond) network architecture. In: *Proc. of the ACM SIGCOMM*. New York: ACM, 2007. 181–192. [doi: 10.1145/1282427.1282402]
- [23] Jacobson V, Smetters D, Thornton J, Plass M, Briggs N, Braynard R. Networking named content. In: *Proc. of the 5th ACM Int'l Conf. on Emerging Networking Experiments and Technologies (CoNEXT)*. New York: ACM, 2009. 1–12. [doi: 10.1145/1658939.1658941]
- [24] Kong LB, Tang SW, Yang DQ, Wang TJ, Gao J. Querying techniques for XML data. *Journal of Software*, 2007,18(6):1400–1418 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/18/1400.htm> [doi: 10.1360/jos181400]
- [25] Lin ZY, Yang DQ, Wang TJ, Zhang DZ. Keyword search over relational databases. *Journal of Software*, 2010,21(10):2454–2476 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3913.htm> [doi: 10.3724/SP.J.1001.2010.03913]
- [26] Cao R, Wu JP, Xu MW. Research on Internet naming. *Journal of Software*, 2009,20(2):363–374 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3389.htm> [doi: 10.3724/SP.J.1001.2009.03389]
- [27] Zhang GQ, Zhang GQ, Yang QF, Cheng SQ, Zhou T. Evolution of the Internet and its cores. *New Journal of Physics*, 2008,10: 123027. [doi: 10.1088/1367-2630/10/12/123027]
- [28] Kaashoek M, Karger D. Koorde: A simple degree-optimal distributed Hash table. In: Kaashoek MF, Stoica I, eds. *Proc. of the 2nd Int'l Workshop on Peer-to-Peer Systems (IPTPS)*. LNCS 2735, Berlin, Heidelberg: Springer-Verlag, 2003. 98–107. [doi: 10.1007/978-3-540-45172-3_9]

附中文参考文献:

- [2] 瞿裕忠, 胡伟, 郑东栋, 仲新宇. 关系数据库模式和本体间映射的研究综述. *计算机研究与发展*, 2008,45(2):300–309.
- [24] 孔令波, 唐世渭, 杨冬青, 王腾蛟, 高军. XML 数据的查询技术. *软件学报*, 2007,18(6):1400–1418. <http://www.jos.org.cn/1000-9825/18/1400.htm> [doi: 10.1360/jos181400]
- [25] 林子雨, 杨冬青, 王腾蛟, 张东. 基于关系数据库的关键词查询. *软件学报*, 2010,21(10):2454–2476. <http://www.jos.org.cn/1000-9825/3913.htm> [doi: 10.3724/SP.J.1001.2010.03913]
- [26] 曹锐, 吴建平, 徐明伟. 互联网命名问题研究. *软件学报*, 2009,20(2):363–374. <http://www.jos.org.cn/1000-9825/3389.htm> [doi: 10.3724/SP.J.1001.2009.03389]



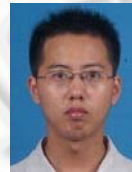
廖祝华(1977—),男,湖南攸县人,博士生,讲师,CCF 会员,主要研究领域为以内容为中心的网络,下一代互联网,网络业务融合.



傅川(1973—),男,高级工程师,CCF 会员,主要研究领域为计算机网络架构,分布式系统,软件工程.



张国清(1965—),男,博士,副研究员,博士生导师,CCF 高级会员,主要研究领域为网络拓扑及应用,网络科学,信息网络,大数据分析.



张国强(1980—),男,博士,副教授,CCF 高级会员,主要研究领域为网络科学,未来网络.



杨景(1952—),男,博士,教授,博士生导师,主要研究领域为网络业务控制和集成体系结构,电信网络管理,互联网产业结构.