

## 网络环境下动态分布内容的自组织语义整合框架\*

廖祝华<sup>1,2</sup>, 张国清<sup>1</sup>, 杨景<sup>1,3</sup>, 边建功<sup>1</sup>, 易爱平<sup>2</sup>

<sup>1</sup>(中国科学院 计算技术研究所, 北京 100190)

<sup>2</sup>(知识处理与网络化制造湖南省普通高校重点实验室(湖南科技大学), 湖南 湘潭 411201)

<sup>3</sup>(中国移动通信研究院, 北京 100053)

通讯作者: 廖祝华, E-mail: liaozhuhua@ict.ac.cn

**摘要:** 随着网络 and 多媒体技术的发展, 新型网络技术不断出现, 并通过互联网络不断融合, 使得网络更加泛在化、开放和动态化。目前, 网络中动态分布的内容爆炸式增长与用户个性化需求之间的矛盾也日益突出。因此, 基于开放的网络环境下海量的内容, 如何提供个性化、智能的内容服务, 已成为研究界和工业界共同关注的问题。基于开放、动态的网络环境, 以个性化的语义整合服务为核心目标, 将语义技术与网络通信机制相结合, 提出了动态分布内容自组织语义整合的网络结构和关键技术的实现机制, 包括关系路由模型、自组织语义整合技术框架和实现流程, 以支持在网络环境下针对动态分布的内容构建个性化和智能的语义整合应用, 并且为动态分布内容的语义整合系统提供工程设计方案和技术实现方法。

**关键词:** 内容互连网络; 关系路由; 自组织语义整合; 通信结构

**中图法分类号:** TP393      **文献标识码:** A

中文引用格式: 廖祝华, 张国清, 杨景, 边建功, 易爱平. 网络环境下动态分布内容的自组织语义整合框架. 软件学报, 2013, 24(12): 2871-2882. <http://www.jos.org.cn/1000-9825/4431.htm>

英文引用格式: Liao ZH, Zhang GQ, Yang J, Bian JG, Yi AP. Self-Organizing semantic integration framework for dynamic and distributed contents in the network environment. Ruan Jian Xue Bao/Journal of Software, 2013, 24(12): 2871-2882 (in Chinese). <http://www.jos.org.cn/1000-9825/4431.htm>

### Self-Organizing Semantic Integration Framework for Dynamic and Distributed Contents in the Network Environment

LIAO Zhu-Hua<sup>1,2</sup>, ZHANG Guo-Qing<sup>1</sup>, YANG Jing<sup>1,3</sup>, BIAN Jian-Gong<sup>1</sup>, YI Ai-Ping<sup>2</sup>

<sup>1</sup>(Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

<sup>2</sup>(Key Laboratory of Knowledge Processing and Networked Manufacturing (Hu'nan University of Science and Technology), Xiangtan 411201, China)

<sup>3</sup>(The Research Institution of China Mobile, Beijing 100053, China)

Corresponding author: LIAO Zhu-Hua, E-mail: liaozhuhua@ict.ac.cn

**Abstract:** With the development of network and multimedia technologies, many new networks are emerging and converging through the Internet, making the network more ubiquitous, open and dynamic. And nowadays the mismatch between the explosive growth of dynamic and distributed content and the personalized needs of users in the network is increasing. Therefore, giving the large-scale contents in the network environment, how to provide personalized, intelligent content services has become the common issues in research community and industry. The purpose of this study is to combine semantic technology and network communication mechanism to support personalized and semantic integration services in the open, dynamic network environment. This paper proposes a self-organizing semantic integration framework with key techniques of realization mechanism for the dynamic, distributed contents, which includes the relational routing

\* 基金项目: 国家科技重大专项(2011ZX03002-004-03); 国家自然科学基金(61174152, 61370227); 湖南省自然科学基金(13JJB004); 北京市自然科学基金(4112057)

收稿时间: 2012-05-21; 修改时间: 2013-01-25; 定稿时间: 2013-05-24

model, the technology framework and implementation process of self-organizing semantic integration. The framework supports building the personalized, intelligent integration applications suited to the dynamic and distributed contents, and provides the engineering design and technical realization methods for semantically integrating the dynamic and distributed contents in network environments.

**Key words:** content networking; relational routing; self-organizing semantic integration; communication architecture

随着互联网中内容的急剧增长,大规模的内容因为存储的优化和用户的操作等因素,使得内容产生各种划分,造成内容碎片化,而碎片化内容的分布使它们之间的语义关联更加复杂并且局部化,进而影响对网络空间中存储内容的评估、甄别、推荐和获取,严重阻碍了网络中内容获取的准确性和完整性.在开放、动态的网络中,目前对海量、分散内容的整合方法很难适用网络用户的个性化需求,其主要原因有:

- (1) 在动态、开放的互联网络中,基于单一、固定网络结构所提出的语义整合方法难以直接应用;
- (2) 当前的网络仅负责端到端二进制数据的传输,缺乏对网络中动态、分布内容的语义认知和整合能力,从而产生了很多无法解决的问题,如:数据传输路径难以动态优化,开放网络中的内容副本难以重用,语义关联的碎片化内容难以按需整合,尤其是难以根据语义整合需求对需要再处理的内容进行个性化的语义重组和高效的数据调度等.

造成这些问题的主要原因有:

- 网络或应用难以解析第三方的内容和语义并及时感知其状态与位置变化;
- 缺乏有效的内容搜索和整合机制;
- 缺乏对第三方内容的智能组织和自适应能力.

因此,要在现有网络架构之上,利用现有的方法实现动态分布内容的智能整合,并且满足 QoS 性能要求,以提供个性化和具有用户体验的内容服务,还具有难以克服的挑战.为此,本文提出基于网络环境的动态分布内容的自组织语义整合方法和系统结构框架,以在网络环境中为用户或上层应用提供个性化和智能的语义整合服务.

## 1 相关工作

面向网络的动态分布内容的语义整合系统与其他整合体系(如 Mashup 应用、Pub/Sub 系统、智能教学系统和语义 Web 应用等)一样,都需要对分布的内容进行不同程度的规范化定义,再按照一定语义要求进行整合.目前,基于开放、分布式环境中不同对象的整合系统或结构,人们主要从 4 方面展开深入研究<sup>[1-3]</sup>,分别是以服务为中心的网络结构,以内容为中心的网络结构,以用户为中心的网络结构和以知识为中心的网络结构.

以服务为中心的网络结构,如基于 SOA 架构的 Web 服务组合<sup>[4]</sup>,其实质上是一种基于集中式注册中心对 Web 服务描述内容的整合.为了充分利用网络的通信资源,目前人们提出了网络资源的虚拟化技术并以服务的形式开放出来.文献[5]提出网络虚拟化技术以对网络层的资源进行抽象,屏蔽实现细节并以服务的方式被外部调用,但目前主要研究的是网络中计算资源、传输资源的虚拟化和开放问题.文献[6]研究对网络中虚拟化资源(如组播)进行服务抽象,从而以服务的方式被外部调用,并可以借鉴 Web 服务组合的方法对虚拟化资源进行组合,完成面向媒体业务的组合.文献[7]针对下一代网络中商业、软件、基础设施与网络的不同层次(尤其是在不同域)的融合管理和治理问题,介绍了多层 SLA 管理方法,以跨层解决这些问题.然而,以服务为中心的网络结构对动态分布的资源命名、查询与功能组合等基本方法还缺乏有效的系统机制和方案,而对于开放、动态网络环境中的语义 Web 服务组合还没有有效的系统模型和方法.

以内容为中心的网络结构,如 CCN/NDN<sup>[8,9]</sup>,ALLOA<sup>[1,10]</sup>等,主要目的是解决网络环境下单个内容获取的通信问题.为此,引入了异步组播和缓存网络化的特征.在 NDN 中,路由器(或解析处理器)通过类似 BGP 的方式向邻居节点宣告并传播内容标识可达性信息,并据此建立基于标识的路由表.用户对内容的请求直接由路由系统依据内容标识路由到内容的存储位置,实现了解析和路由的统一.相比于 CCN/NDN 从获取单个内容块出发,欧洲未来内容网络组(FCN)提出的 ALLOA(autonomic layer-less object architecture)结构则从多媒体信息获取出发,提出了复杂内容对象(content object)的概念,将互联网络中的内容对象看成是一个多态或功能整体性的容

器,可以包含媒体的内容、操作规范、对外的行为影响、与其他对象的关系以及对象的属性描述等.内容对象可以智能地组合、分解甚至能够智能地路由和转换等<sup>[10]</sup>.从总体来看,以内容为中心的网络架构研究为动态分布内容的语义整合提供了基础和方法,但当前还处于刚起步阶段.本文提出的自组织语义整合系统结构是在其思路和机理上的改造和扩展,并以面向动态分布内容的语义整合作为新的目标.

以用户为中心的网络结构主要基于用户的特征、爱好和行为等建立用户模型<sup>[11]</sup>,然后通过用户群体特征对用户进行内容过滤和推荐,而对内容的路由查找和获取仍然采用传统方法.例如,当前社交网络、博客和微博等得到了广泛的应用,然而其应用内容的整合方式主要是依靠参与者关系网络来进行传播、分类和推荐<sup>[12]</sup>,并不具备高度智能和语义的整合能力,与动态分布内容的自组织语义整合方法完全迥异.

以知识为中心的网络结构目前主要研究基于 Web 技术的基础上构建语义网络、知识网格(knowledge grid)<sup>[3]</sup>,它们关注的主要是内容的语义处理,并不考虑通信问题.认知网络(cognitive network)<sup>[13]</sup>虽然考虑在通信网络中建立知识平面<sup>[2]</sup>,但其目标是增强网络对各种通信资源的认知和调度能力.

因此,这 4 种针对不同对象的整合网络架构还不能解决网络环境下动态分布内容的语义整合问题,迫切需要研究和设计新的方法和体系架构.

## 2 自组织语义整合及通信结构

### 2.1 总体结构

整合有集成和再组织之意.为了便于在开放的网络环境中对各种内容进行语义整合,内容除了指交付给用户的原始内容(如文字、表格以及各种富媒体内容如 flash 视频、3D 图形等)之外,还包括描述内容的 meta 信息,如描述内容的属性与语义关联关系等.因此,开放网络环境中的内容对象可以抽象定义为  $D(P,r,b)$ ,即包括属性  $p$ 、操作规则  $r$  和对外表现的行为  $b$ .

对网络中动态分布内容的处理需要包括内容的语义查询、语义整合和转发处理.相应地,我们将分析这 3 种类型的处理,并对它们进行统一组织:

第 1 种是语义查询处理,定义为  $M_q \rightarrow \frac{Q(M_q)}{D} = D_o$ ,即根据外部查询命令  $M_q$  对系统本身拥有的内容集  $D$  进行查询、过滤和排序等处理,然后输出与需求最相关的内容;

第 2 种是语义整合处理,定义为  $M_I \rightarrow \frac{I(M_I)}{D} = D_o$ ,即根据外部整合命令  $M_I$  对系统拥有的内容集  $D$  进行语义整合、抑制冗余和数据封装等处理,然后输出与整合需求相关的组合内容;

第 3 种是内容转发处理,定义为  $H_{u,s} \rightarrow \Phi(D_I) = D_I$ ,即当数据所在地与用户之间存在距离时(转发次数不为 0),则对输入的内容  $D_I$  按需要进行单播、组播等转发方式的处理,使得内容有效到达目的地,但内容本身的数据和结构不发生变化.

开放内容源上的语义查询处理和语义整合处理可以看成是网络节点上的计算型语义处理,因此,语义整合实际上是根据用户请求从网络中发现相关的内容,然后在网络中聚集这些内容,并对其进行按需组织和加工(包括格式转换和语义整合等),最后返回给用户的过程.总之,基于自组织的语义整合方式是一种对动态、分布的内容在网络范围内进行自动协商并给出组织方案的新型语义整合服务模式.假设用户的整合需求为  $I_u$ ,返回的整合结果为  $D_u$ ,则整个基于网络环境的自组织语义整合方式可以抽象地表达为

$$I_u \rightarrow \sum_{i=1}^n \left( M_i^i \left( \sum_{j=1}^m M_q^j / \Phi(D_o^j) \right) / \Phi(D_o^i) \right) = D_u$$

其自组织语义整合及其通信结构可以抽象地表示为图 1 所示的结构.

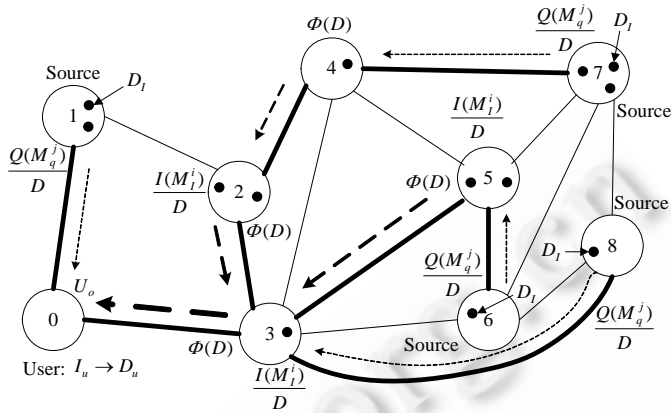


Fig.1 Self-Organizing semantic integration and its communication structure

图 1 自组织语义整合及其通信结构

2.2 系统体系结构

为了基于自组织语义整合的方法和机制实现网络中动态分布内容的语义整合,我们设计了自组织语义整合体系架构,其层次划分如图 2 所示.

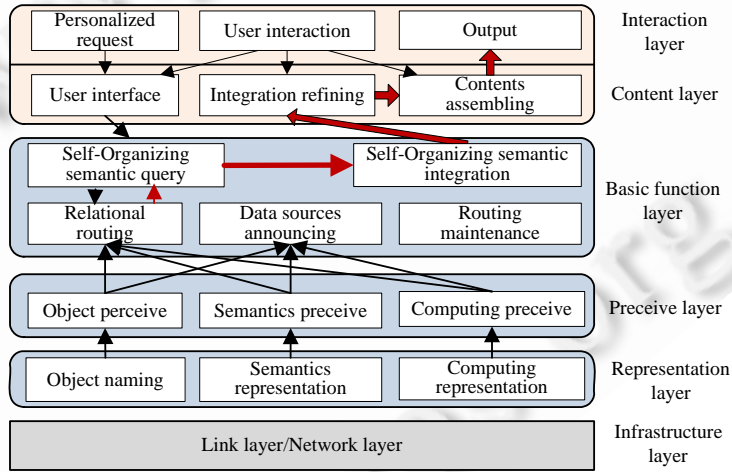


Fig.2 Self-Organizing semantic integration system framework

图 2 自组织语义整合体系框架

每一层实现不同的功能并能被上层调用:

- (1) 设施层:传统网络提供的网络硬件设施以及网络链路层或网络层的通信协议;
- (2) 表示层:采用标准的语义表示方法负责对网络范围内各地内容资源以及语义信息进行形式化表示,异构的内容源具有相互的语义映射机制;另外,对各路由节点的计算能力表示包括语义查询、语义整合的功能声明以及计算性能的量化;
- (3) 认知层:实现对分布的命名内容和语义的认知以及主要业务整合行为和应答结果的认知;
- (4) 基本功能层:实现网络中动态、分布的命名内容的语义查询和语义整合.根据具体语义要求,对相关局部内容的语义关系进行汇聚和整合,在数据传输过程中,负责路由和转发以及路由的维护等;
- (5) 内容层:具体由各路由节点合作实现.主要根据用户的语义请求调用下层的基本功能,对下层提供的

逻辑候选序列进行个性化整合和改造等,实现内容关联逻辑合成,并根据关联逻辑汇聚分散的最近内容对象副本完成个性化的语义整合;

- (6) 交互层:负责个性化语义整合请求的提交、自组织语义整合过程中的参数设置、调节以及交互控制接口等。

其中,第1层是属于传统网络范畴,主要提供路由节点邻居之间的数据通信能力;中间的第2层~第5层是自组织语义整合的核心层,主要采用规范化和协议的描述方法以及自组织的语义查询和整合方法实现对自组织语义整合系统中基本要素的认知、控制和内容的语义整合,对获取的局部内容关联逻辑进行全局再组织,从而按照分布的语义整合流程汇聚用户需要的内容;而第6层属于应用层,主要负责对核心层的功能调用、用户复杂请求的形式化,并完成自组织语义整合过程中的调用和交互等。

该框架中关键部分是关系路由、语义整合和分布式控制的机制,主要包括:

- (1) 内容的表示与感知,即动态、开放的网络环境中分布内容的规范表示与语义查询;
- (2) 局部关联逻辑的认知,完成单个节点中内容语义的查询和整合等;
- (3) 关系路由,即根据用户或路由节点请求中的参照对象或者层次分类前缀之间及与要获取的内容之间的关系对请求进行路由和转发;
- (4) 语义查询和整合,即根据用户请求中的语义对分布、动态的内容进行全局的语义查询,并按用户要求进行对应模式的语义整合.语义整合中需要考虑用户的个性化需求以及整合的效率等重要指标;
- (5) 路由维护,即在自组织语义整合过程中考虑请求转发和结果返回的路由维护、路由重寻,以保证自组织语义整合在动态、开放网络中可靠、高效地完成。

### 2.3 关系路由模型

在动态分布的网络环境中,自组织的语义整合方式能够解决内容服务的个性化问题,解决持续满足用户需求的问题,解决系统复杂化、环境变化所带来的语义整合问题.而研究高效率的自组织路由机制是网络中动态分布内容的语义整合的关键.互联网络中数据传输协议和路由算法是当前受到最广泛研究的问题,在自治网络中不存在缺省的数据传输路由,网络中的每个节点都要求具有独立查询并传输数据的能力.因此,可把整个网络看成一个自治系统,网络中的节点看成路由器节点。

内容层次化的划分,可使得大粒度的内容集合容易划分成为多个子集,并可以映射为网络空间中的不同分布区域;其次,标准的命名使得内容名字之间的关系能够逻辑关联网络全局范围的内容对象和副本.为使网络能够认知动态分布的内容,提出了一种能够实现动态语义导航的关系路由机制,其结构如图3所示,即在内容互联网络之上,通过路由节点(本文我们又称其为 *guider* 节点)与数据源互连建立一个动态层次的语义导航网络,从而可以使得网络能够按用户的语义需求识别和查询动态分布的内容,即基于参考信息进行路由导航,直到在数据源或内容缓存节点根据查询约束完成两种基本的查询方式:属性约束查询和关系约束查询.因此,用户可以使用层次化的参考信息(*reference*)以及约束(*constraint*)来查询和聚集网络中动态分布的内容,见文献[14].

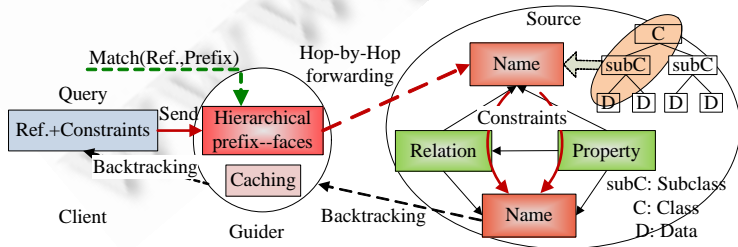


Fig.3 Relational routing model

图3 关系路由模型

### 3 语义整合流程

本节将探讨两类动态分布内容的语义整合方式,分别是简单整合和复杂整合.其中,简单整合是基于关联或属性约束;复杂整合则提供各种个性化语义表达方式,从而可以完成比较复杂和功能较强的动态分布的语义整合.下面分别介绍简单和复杂整合这两种基本形式的语义整合流程,再归纳语义整合的总体流程.

#### 3.1 简单整合业务流程

在语义整合中,个性化的用户查询的流程如图 4 所示.

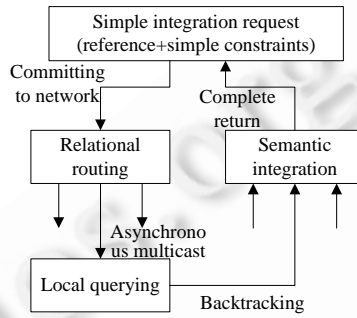


Fig.4 Simple integration process

图 4 简单整合业务流程

总体流程是:首先,用户将形式化的查询请求进行网络提交,其中查询格式具有以下形式:

“参考对象”+“属性约束”和/或“关系约束”.

网络根据查询请求中的参考对象调用关系路由机制,将查询请求转发(异步组播)到所有潜在的数据源中.然后,在数据源中根据查询约束进行局部查询,局部查询的结果沿着请求的转发路径原路返回.在原路返回的过程中,路由节点对结果进行收集,过滤重复的内容,并返回尽可能完备的结果到用户端.整个完备返回的保证机制是由关系路由机制和路由节点的收集处理联合保证.

对于简单整合,如用户输入格式为“ $a\_R(r(a,?))$ ”的字符串描述其整合需求,可以直接用关系路由模型就可以实现.其整合过程与关联关系查询相同,不同点是关联关系查询中的关联对象可以是分类,而整合请求中内容参考对象  $a$  一般是比较确定的内容对象,“ $R(r(a,?))$ ”表示关系约束,即  $a$  与要获取的内容之间存在  $R$  关联关系.因此,我们也把这种整合称为 Forward 整合.详细的 Forward 整合过程如图 5 所示:当整合请求通过 guider 节点转发到潜在的数据源(如 source1,source2)时,将会在本地执行查询(基于关联关系查询),然后原路返回各个结果(如  $r(a,b1),r(a,b2)$ ).在原路返回的过程中,guider 节点将对来自不同节点的返回结果进行合并,直抵用户端后,将最终整合结果呈现给用户.

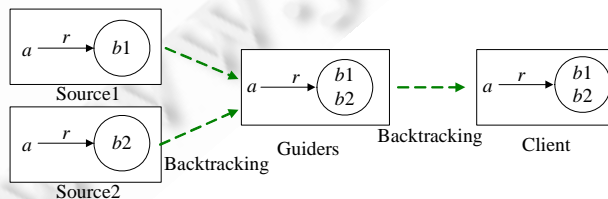


Fig.5 Self-Organizing process of simple (forward) integration

图 5 简单(forward)整合的自组织处理过程

#### 3.2 复杂整合业务流程

在语义整合中,为完成用户复杂的整合目标,需要对动态分布的内容进行复杂整合,其具体业务流程如图 6

所示.

用户首先提交个性化的语义整合目标(形式化为参考对象+复杂整合目标),然后进行网络提交.网络首先根据参考对象并在对应整合模式算法的指导下智能调用关系路由机制,将用户请求转发(异步组播)到所有潜在的数据源中进行局部内容的关联整合;然后,局部整合的结果原路返回,在多个结果汇聚的路由节点逐步多次完成复杂模式的语义整合并对冗余结果进行过滤;最后,由关系路由机制以及对应整合模式算法保证全局整合的结果的完备返回,完成用户的复杂整合业务目标.

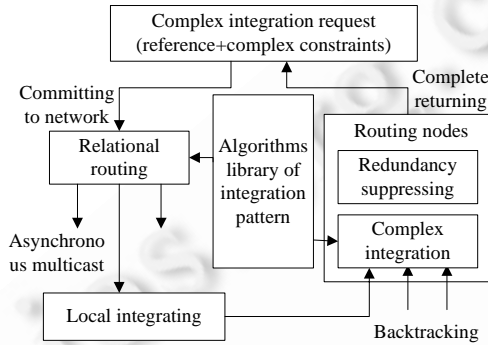


Fig.6 Complex integration process

图 6 复杂整合业务流程

下面介绍两种较为复杂的整合方式:

(1) 接合(joint)整合,即用户给出整合请求的名字为“ $b\_R\langle Joint(r(?),b),s(b,?) \rangle$ ”格式的语义整合.这是一种有约束的简单整合及其变种的结合,约束条件即前驱的后一内容对象和后继的前一内容对象必须相同.

详细的接合整合过程如图 7 所示.当接合整合请求通过 *guider* 节点转发到潜在的数据源中时,具体的接合整合请求将被本地执行查询和初步接合整合.然后,整合结果原路返回,并在 *guider* 节点中通过接合整合算法对各数据源的整合结果进行再次语义整合(如 *source1* 返回  $r\text{-}s(a1,b,c1)$ ,*source2* 返回  $r\text{-}s(a2,b,c2)$ ,*source3* 返回  $r(a3,b)$ ,*source4* 返回  $s(b,c3)$ ),直到返回到用户端;最终,用户将获取接合整合的内容序列(如 $(a1,b,c1)$ , $(a1,b,c2)$ 等).

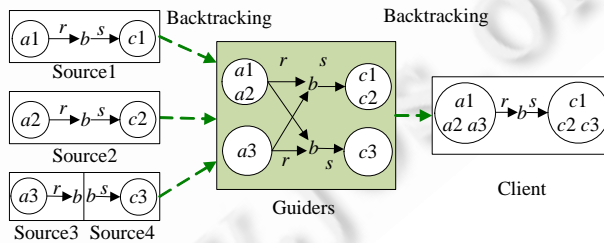


Fig.7 Self-Organizing process of joint integration

图 7 接合整合的自组织处理过程

(2) 桥接(bridging)整合,即用户给出整合请求的名字为“ $a\_R\langle Bridging(r(a,?),s(?,c)) \rangle$ ”格式或者“ $c\_R\langle Bridging(r(a,?),s(?,c)) \rangle$ ”格式的语义整合.这也是有约束的简单整合及其变种的结合,约束条件即指定前驱的前一内容对象和后继的后一内容对象,使得某一个内容对象能够同时满足与这两个内容对象的对应关联关系.

详细的桥接整合过程如图 8 所示.当桥接整合请求通过 *guider* 节点转发到潜在的数据源中时,具体的语义整合请求首先将在本地执行查询,对于初步查询得到的结果首先在数据源端进行初步整合(如 *source1* 的初步整合结果为“ $(a,b1) \cap (b1,c)$ ”,*source2* 的初步整合结果为“ $(a,b2) \cap (b2,c)$ ”).而对于有些数据源端的查询结果比较简单,不需语义整合,直接原路返回(如 *source3* 和 *source4*).在返回的过程中,中间 *guider* 节点根据桥接整合算法不断对来自不同数据源的结果进行再次整合,直到最后到达用户端.最终,用户将获取多个需要的内容序列(如

(a,b1,c),(a,b2,c),等等).

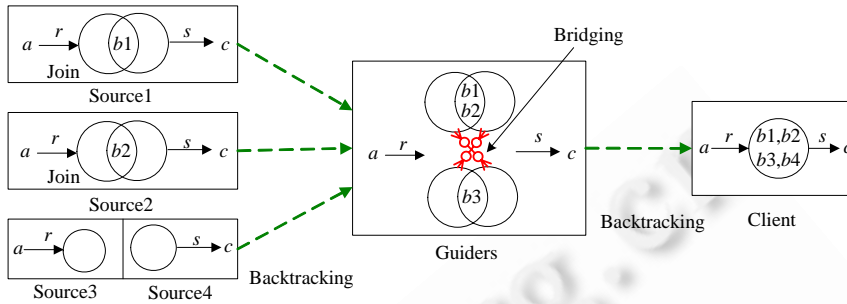


Fig.8 Self-Organizing process of bridging integration

图8 桥接整合的自组织处理过程

(3) 复杂组合整合,如用户可以指定序列中的若干关键内容以及关键内容与即将整合的未知内容的关联关系,而且不同关键内容与未知内容的连接关系可以不相同.例如,对于如下复杂组合整合请求:

$$a\_R(\text{Bridging}(r1(a,?),r2(?),c)),\text{Joint}(d=r4(?),e),r5(e,?)),r3(c,d)).$$

这种整合方式最终在用户端生成的是连接给定关键节点的多条有向路径图.详细的复杂组合整合过程如图9所示.当整合请求通过 *guider* 节点转发到潜在的数据源时,首先根据复杂组合整合的约束条件(包括简单整合、接合整合和桥接整合的组合约束)在数据源端多次执行查询,然后将查询结果根据复杂组合整合各部分的整合要求将其整合成全部或部分满足需求的内容子序列,然后原路返回.在返回的过程中,*guider* 节点会再次根据复杂组合整合算法将对来自不同节点的内容子序列进行多次整合.最终,当数据传输到用户端时,用户可以获取满足复杂组合整合的所有约束条件的多个内容序列.

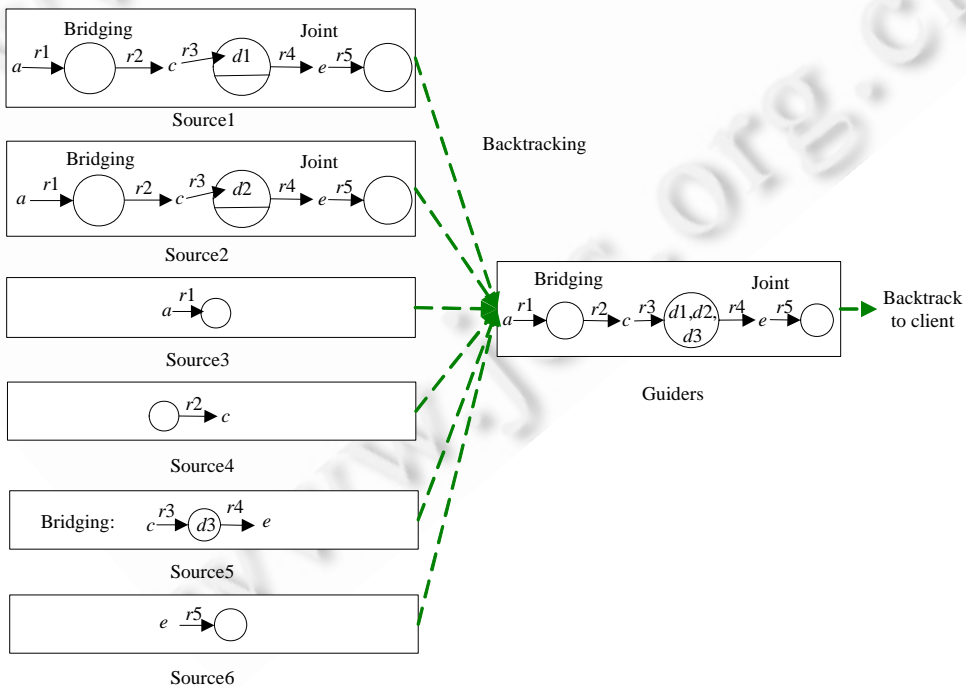


Fig.9 Self-Organizing process of complex compounding integration

图9 复杂组合整合的自组织处理过程



### 3.3 总体语义整合流程

如图 10 所示,语义整合的总体技术流程是:首先理解用户的个性化目标并转换成紧凑、形式化的业务生成请求,发送给网络(由路由节点转发).网络根据用户请求采用关系路由方法,根据参考对象和索引,将查询请求转发到潜在的数据源端,然后,网络根据用户请求对分布的内容使各节点单元协作完成语义查询,并从语义逻辑上进行整合,生成满足用户需求的内容关联逻辑.整合过程中调用内容复杂整合机制,将分布的相关内容按用户需求进行关联,调用路由维护机制保证在动态网络中内容关联逻辑和业务内容的原路返回,从而在用户端获取用户需要的分布在动态网络中的最优内容集合.为了满足自组织整合过程中的流量工程优化目标,可设置专门的高层控制机构,例如,通过控制整合过程中的请求转发速度、数量,对结果返回排序、冗余抑制等方法来控制数据包的线速转发、自组织语义整合的负载均衡和数据传输过程中的流量优化.为了适应网络环境的动态变化,可设置路由维护模块,专门对失效节点进行探测,失效边缘节点进行数据源更新和请求续发等,以维护数据源通告路由和请求转发路由.另外,用户可以在内容自组织整合过程中实行相关参数的调节,例如:整合请求的参考信息、约束条件的更改,整合请求在路由节点中的逗留和撤销时间,整合中进行高层优化控制的相关参数以及响应结果的返回策略用以兼顾响应速度或优化传输流量等目标.因此,整个动态分布内容的自组织整合系统是一个闭环系统,从而达到最优调度相关网络资源对动态分布内容进行事务性的整合处理,协调各种网络设备和路由资源建立定向的语义路由和内容分发通道以及自组织整合的反馈可控制机制.

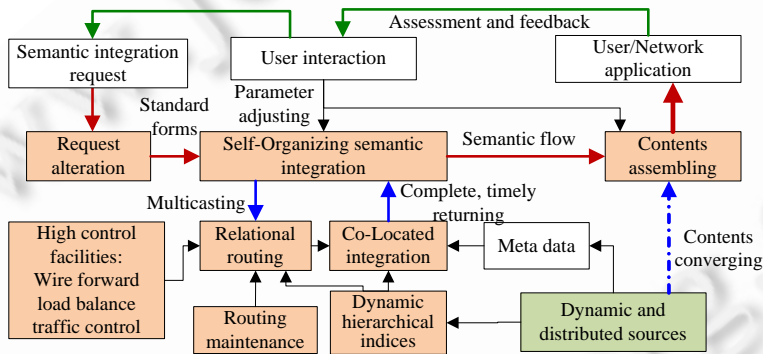


Fig.10 General processes of semantic integration  
图 10 语义整合总体流程

## 4 系统实现和测试

本文提出的自组织语义整合框架可以在网络层或者应用层实现,只要下层的网络设施(如链路层)能提供相邻节点之间的通信即可.为了验证自组织语义整合机制的性能和效率,我们在应用层,基于 CCN/NDN 协议<sup>[8]</sup>,采用 Java 语言编程实现了自组织语义整合系统原型,并在局域网中进行了测试.建立了一个由 1 020 个路由节点组成,分布在 68 台物理计算机上的 Overlay 网络,每台机器平均部署 15 个路由节点,另外设置 1 054 个数据源与路由节点随机相连.计算机硬件配置是 Intel Pentium 4 CPU 和 512M~2G 的内存组成,安装 Windows 7 操作系统和 Java 开发套件 JDK1.6.在各数据源中,共分布有计算机、通信和电子信息专业方面的 30 门多媒体课程内容之间的关联数据(由 MySQL 数据库存储).多媒体课程内容采用层次化命名,主要由学科名、课程名、章、节以及多媒体内容的名称组成.另外还定义了多媒体内容的属性和内容之间的关联关系,关联关系包括显式的关联,如课程编排顺序、包含关系、交叉关系、共用关系和并列关系等;隐式关联主要是由显式关联隐含给出.数据库中的记录指出了内容之间的显式关联,每个数据源的数据记录平均约为 3 000 条.

我们首先考察了网络中分布的参考对象副本在不同规模下语义整合的效率,为了正确地统计总的整合时间,我们首先选择参考对象副本的平均规模是 3,然后我们对每种整合方式分别重复测试 10 次.整合时间(也可以说是总的路由时间)是指从整合请求发送出去到结果中最后一个数据包到达的时间总和.图 11(a)显示了简单整

合(forward integration)、复杂整合(joint integration 和 bridging integration)方式的整合时间与路由跳数的关系. 结果中第 1 个数据包的到达时间(如图中柱形黑色部分最高点)和最后一个数据包的到达时间(如图中柱形最高点)随路由跳数大小的变化. 结果显示:随着路由跳数逐步从 1~8 变化,简单整合与复杂整合的整合时间均呈亚线性增加,并且所有整合方式都能在较短的时间内完成;但整合方式越复杂,其响应结果持续到达时间则逐步延长.

图 11(b)显示了 3 种整合方式的整合时间与副本规模的关系.为了正确地统计整合时间,我们首先选择用户连入网络的节点,使得用户到各数据源的平均跳数是 6;然后,我们对每种整合方式分别重复测试 10 次.结果中第 1 个数据包的到达时间和最后一个数据包的到达时间随副本规模的变化. 结果显示:随着参考对象副本规模逐步从 2 增长到 16,响应结果持续到达时间和总的整合时间开销亚线性增加.但各动态分布内容的语义整合可以在比较短的时间之内完成.语义整合能在较短时间完成的原因是因为各数据源的分布性特征,并且用户查询转发和应答结果返回采用了异步组播机制,因此查询和结果返回都能并发进行,而且本地化副本可减少组播分支与距离.

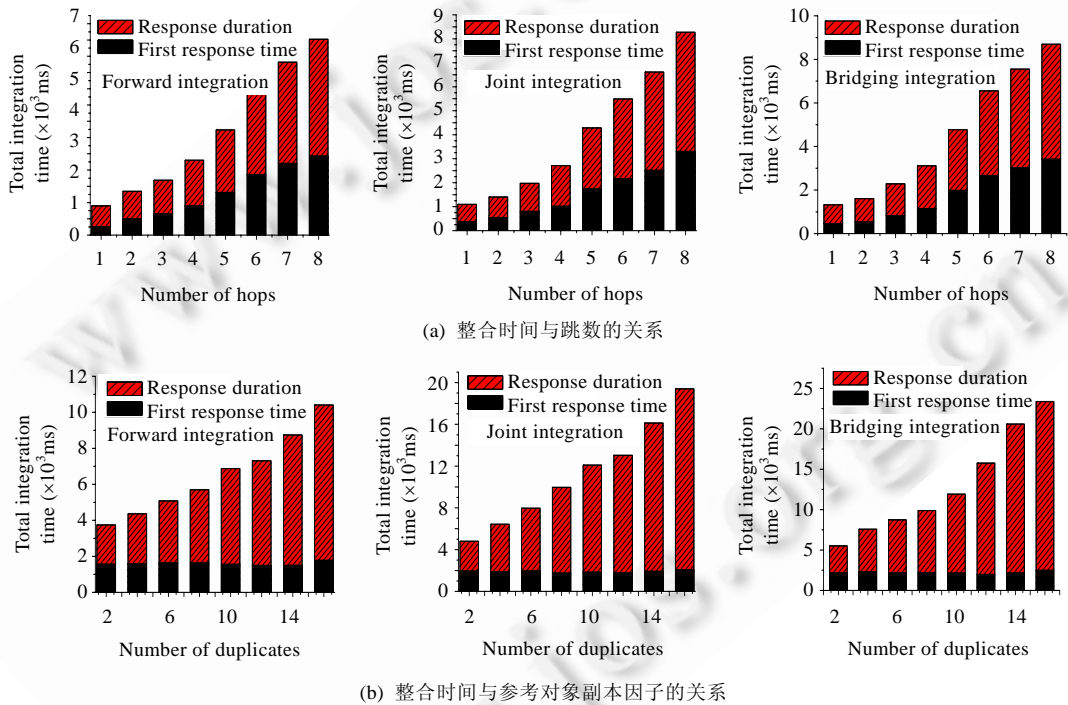


Fig.11 Correlation between the efficiency of semantic integration and forward hops/duplicate factor

图 11 语义整合效率分别与跳数、副本因子的关系

另外,我们测试了复杂组合整合方式的整合效率.图 12(a)显示:当转发次数固定为 6 时,候选序列数成倍增长,而整合的时间则以较低的速率增长.这是由于数据源的分布,语义整合请求的路由和应答结果的返回都可以并发进行,而内容的整合从源端就开始直到用户端不断以级联的方式逐步整合和收敛.

同样,如图 12(b)所示:候选序列长度成倍增长,则语义整合的时间也只以较低的速率增加.这是因为数据源的分布性,使得内容的语义整合计算比较分散,从而并发操作的规模更大,使得语义整合的计算量并不是随着序列长度的增长而线性增加.

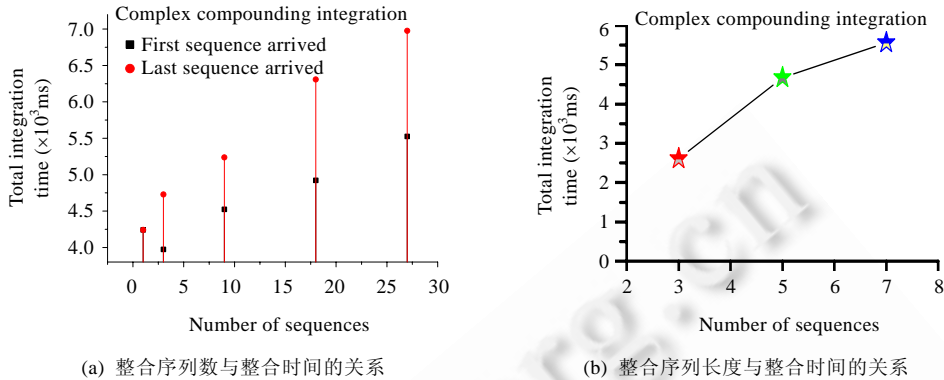


Fig.12 Efficiency measurement of complex compounding integration

图 12 复杂组合整合方式的效率测试

## 5 结 论

本文基于开放的网络环境,以动态分布内容的个性化语义整合为核心目标,提出了面向内容自组织语义整合的系统框架和关键技术的实现机制,并将语义技术与网络通信机制结合.由于继承了内容中心网络中对数据的层次命名及其位置无关性特点,并增强了路由节点的自主语义处理能力,本网络架构支持网络环境下数据查询、语义整合与位置无关,具有动态自适应的语义整合能力、灵活与智能的数据处理能力以及节点间的自组织能力,因而能够实现动态分布数据的语义整合与获取,提高动态分布数据整合的准确性和个性化.为在网络中构建个性化和智能的内容的语义整合应用,对自组织的语义整合方法以及实现机制、框架和流程进行了重点研究和介绍,为动态分布内容的语义整合系统提供了工程设计方案和技术实现方法.

本系统在请求和结果返回的路由和转发方面,采用了基于标识的关系路由机制,使得动态分布内容的语义整合既可以在覆盖网层实现,也可以在网络层实现,并保证了系统路由层遵守“细腰”原则.对于请求转发和应答结果返回都采用了异步组播方式,从而语义整合和路由功能是由若干相关节点协作完成,而在各相关节点中语义整合实现机制并不太复杂.虽然简单语义增强仍会对网络的性能造成影响,但在系统中考虑了使用流量控制、负载均衡以及线速转发等方面的优化控制策略,从而满足网络流量工程的要求,具有可部署和可扩展性的潜力,并能适用于移动、自组织的网络环境.在功能上,本系统可以对动态分布内容实现语义整合,并且可以对命名内容进行不同方式的复杂整合,从而用户或应用可以根据自己的意愿,应用不同形式的语义整合方式在全网网络范围内获得个性化、全局最优和语义完整的内容.

## References:

- [1] Zahariadis T, Daras P, Bouwen J, Niebert N, Griffin D, Alvarez F, Camarillo G. Towards a content-centric Internet. In: Proc. of the Towards the Future Internet—A European Research Perspective. IOS Press, 2010. 227–236. [doi: 10.3233/978-1-60750-539-6-227]
- [2] Clark DD, Partridge C, Ramming JC, Wroclawski JT. A knowledge plane for the Internet. In: Proc. of the Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM). New York: ACM Press, 2003. 3–10. [doi: 10.1145/863955.863957]
- [3] Cannataro M, Talia D. Semantics and knowledge grids: Building the next-generation grid. IEEE Intelligent Systems, 2004,19(1): 56–63. [doi: 10.1109/MIS.2004.1265886]
- [4] Krummenacher R, Domingue J, Pedrinaci C, Simperl E. SOA4All: Towards a global service delivery platform. In: Proc. of the Towards the Future Internet. IOS Press, 2010. 61–172. [doi: 10.3233/978-1-60750-539-6-161]
- [5] Zhang Y, Sun ZG. Virtualization technology for trustworthy network research. Chinese Journal of Computers, 2009,32(3):417–423 (in Chinese with English abstract).

- [6] Kim JW, Han SW, Yi DH, Kim NG, Kuo CCJ. Media-Oriented service composition with service overlay networks. *Journal of Communications*, 2010,5(5):374–389. [doi: 10.4304/jcm.5.5.374-389]
- [7] Theilmann W, Baresi L. Multi-Level SLAs for harmonized management in the future Internet. In: *Proc. of the Towards the Future Internet—A European Research Perspective*. 2009. 193–202. [doi: 10.3233/978-1-60750-007-0-193]
- [8] Jacobson V, Smetters DK, Thornton JD, Plass M, Briggs N. Networking named content. In: *Proc. of the 5th Int'l Conf. on Emerging Networking Experiments and Technologies (ACM CoNEXT)*. Rome, 2009. [doi: 10.1145/1658939.1658941]
- [9] Zhang LX, Estrin D, Burke J, Jacobson V, Thornto JD, Smetters DK, Zhang BC, Tsudik G, Kriokov D, Massey D, Papadopoale C, Abdelzaher T, Wang L, Crowley P, Yeh E. Named data networking (NDN) project. PARC Technical Report, 2010-003, 2010. <http://www.named-data.net/>
- [10] Alduán M, Álvarez F, Zahariadis T, Nikolakis N, Chatzipapadopoulos F, Jiménez D, Menéndez JM. Architectures for future media Internet. In: *Proc. of the Int'l Conf. on User Centric Media (UCMEDIA 2010)*. LNICST 60, 2012. 105–112. [doi: 10.1007/978-3-642-35145-7\_14]
- [11] Kobsa A. Generic user modeling systems. In: *Proc. of the Adaptive Web*. LNCS 4321, 2007. 136–154. [doi: 10.1007/978-3-540-72079-9\_4]
- [12] Pessemier T, Deryckere T, Martens L. Context aware recommendations for user-generated content on a social network site. In: *Proc. of the EuroITV 2009*. Leuven, 2009. [doi: 10.1145/1542084.1542108]
- [13] Thomas RW, Friend DH, DeSilva LA, MacKenzie AB, Tech V. Cognitive networks: Adaptation and learning to achieve end-to-end performance objectives. *IEEE Communications Magazine*, 2006,44(12):51–57. [doi: 10.1109/MCOM.2006.273099]
- [14] Liao ZH, Zhang GQ, Yang J, Fu C, Zhang GQ. The scheme of relational routing for querying semantic media on Internetwork. *Ruan Jian Xue Bao/Journal of Software*, 2012,23(10):2760–2771 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4176.htm> [doi: 10.3724/SP.J.1001.2012.04176]

#### 附中文参考文献:

- [5] 张怡,孙志刚.面向可信网络研究的虚拟化技术. *计算机学报*,2009,32(3):417–423.
- [14] 廖祝华,张国清,杨景,傅川,张国强.互网络中语义媒体查询的关系路由方法. *软件学报*,2012,23(10):2760–2771. <http://www.jos.org.cn/1000-9825/4176.htm> [doi: 10.3724/SP.J.1001.2012.04176]



廖祝华(1977—),男,湖南攸县人,博士,讲师,主要研究领域为移动计算,以内容为中心的网络,网络数据获取与整合.

E-mail: liaozhuhua@ict.ac.cn



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

E-mail: gqzhang@ict.ac.cn



边建功(1987—),男,硕士生,主要研究领域为网络数据分析,在线社会网.

E-mail: bianjianguo@software.ict.ac.cn



杨景(1952—),男,研究员,博士生导师,主要研究领域为网络业务控制,集成体系结构,电信网络管理.

E-mail: yangjingyi@chinamobile.com



易爱平(1979—),女,硕士生,主要研究领域为内容中心网络,分布式知识处理.

E-mail: apyi@hnust.edu.cn