ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.17, No.11, November 2006, pp.2255−2263
DOI: 10.1360/jos172255

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

信息网格中基于本体的 Web 服务动态集成和重构*

陈 磊[1+], 韩 颖[2], 李三立[1]

[1](清华大学 计算机科学与技术系,北京 100084)

[2](北京大学 中国语言文学系,北京 100871)

Dynamic Integration and Construct of Web Services Based on Ontology in Information Grid

CHEN Lei[1+], HAN Ying[2], LI San-Li[1]

[1](Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

[2](Department of Chinese Language and Literature, Peking University, Beijing 100871, China)

+ Corresponding author: Phn: +86-10-62775830, E-mail: c-L03@mails.tsinghua.edu.cn, http://www.tsinghua.edu.cn

Chen L, Han Y, Li SL. Dynamic integration and construct of Web services based on ontology in information grid. *Journal of Software*, 2006,17(11):2255−2263. http://www.jos.org.cn/1000-9825/17/2255.htm

Abstract: Web services management and organizing technologies based on syntax level can not meet the need of getting relevant information, responding to diversity service request and reusing Web services in information grid . In this paper, a Web service dynamic integrating and constructing (WS-DIC) strategy based on ontology for information grid is proposed. By the ontology and its reasoning capability, WS-DIC can reuse existing Web services to generate an optimized integrating or constructing paths set to meet the need of request diversity and information relevancy. By the abstraction and formalized model of WS-DIC, the regulation and arithmetic of dynamic integration and reconstruction is discussed. Experimenters indicate that WS-DIC can effectively get the optimized integrating or constructing path and get more advantage then full-text search and SQL query.

Key words: Web service; ontology; information grid; dynamic; integration

摘 要: 基于语法的 Web 服务资源组织和管理策略不能满足信息网格中服务请求多样性和信息关联的需要.提出了一种基于本体的 Web 服务动态集成和重构策略(Web service dynamic integrating and constructing,简称 WS-DIC).该策略以本体及其推理能力为核心,复用信息网格中已有服务,生成优化重构路径集合完成服务集成,满足请求多样性和信息关联的需要.通过对该策略的抽象和形式化描述,讨论了动态重构和集成规则,并设计了动态重构和集成算法.模拟实验表明,与传统的全文检索和数据库查询方式相比,该策略可以按照用户请求,通过服务重构集成,满足请求多样性并准确、全面地获取关联信息.

关键词: Web 服务;本体;信息网格;动态;集成

中图法分类号: TP393 文献标识码: A

信息网格是以消除信息孤岛,实现信息共享、管理和服务的系统,其研究的核心问题是信息共享[1].信息网

* Supported by the SEC e-Institute: Shanghai High Institutions Grid Project of China (上海高校网格 e-研究院资助项目); the 211 Project of China (211 工程项目)

Received 2006-06-09; Accepted 2006-08-07

格中信息格式多样、异构性强、信息量大、信息内容动态变化和信息源分布自治等特点[2],决定了采用传统“仓储式”信息管理方法难以满足网格信息共享的要求.以 Web 服务的方式包装信息源,适用于信息源数目多、信息量大、各局部信息源自治性很高、异构性强且局部信息经常动态变化的网格信息环境.

Web 服务技术吸收了分布式计算、Grid 计算和 XML 等各种技术的优点,通过采用 WSDL(Web services description language),UDDI(universal description, discovery and integration)和 SOAP(simple object access protocol)等基于 XML 的标准和协议,解决了异构分布式计算以及代码与数据重用等问题,具有高度的互操作性、跨平台性和松耦合的特点.考虑信息网格中服务请求多样性和信息关联性,需要大量服务资源.近年来,对信息网格中的 Web 服务资源的组织与管理成为研究热点.目前,对于 Web 服务资源组织与管理的研究很多.已有的网格资源组织模型主要包括:(1) 集中式的资源管理方法,如 Globus 计算网格中的 MDS(metacomputing directory service)[3]实现了基于 LDAP(light directory access protocol)的树状元数据目录服务;Condor[4]实现了不依赖全局资源命名,而依靠属性匹配的集中式的资源共享系统;(2) 采用资源路由机制[5].这种思想借鉴了当前 Internet 的 IP 路由的成功机制;(3) 资源空间模型结构,如 Rajasekar 等人提出的元数据资源空间模型[6].

但上述 Web 服务资源组织与管理技术对服务的描述仅限于语法层次,不能表达语义信息,因此不能满足服务请求多样性和信息关联的需要.基于语义的服务资源组织部分地解决了信息网格中 Web 服务的重构问题,比较典型的研究成果包括基于 XML 工作流描述的 eFlow[7]系统、DynFlow[8,9];基于 DAML-S(DAML(DARPA Agent markup language) service)[10],Agent 的服务自动组合技术[11,12];基于 DAML-S,SHOP2[13]的服务组合方法[14];基于 OWL(Web ontology language)[15]的 BSCM(best-result service composition method)[16]和 DOSCM (domain ontology service composition method)[17]等.上述研究成果大多关注服务组合的成功率和组合效率,对服务满意度的评价方法也大多集中在服务结果的准确度方面.但在信息网格中,由于信息的多样性和关联性,还需要考虑如何在满足一定精确度的条件下获取更全面的信息.本文基于本体对领域内概念与概念间关系的精确描述,提出了一种 Web 服务动态集成和重构策略(Web service dynamic integrating and constructing,简称 WS-DIC),该策略利用本体及其推理能力,复用已有服务,生成优化的集成和重构路径,以解决信息网格中服务请求多样性和信息关联的问题.

本文首先依据抽象代数理论,参考 W3C 的最新推荐标准 OWL,形式化描述 WS-DIC 策略,并在相关命题证明的基础上讨论动态集成和重构算法.第 2 节采用比较实验的方法验证策略的有效性.最后总结全文并介绍 WS-DIC 的完善计划以及未来在上海医学网格上的实现计划.

1 WS-DIC 策略

1.1 基于本体的Web服务和服务请求

领域本体是对特定领域内概念及概念间关系的精确描述.领域本体可以用五元组表示:$O=(C,R,Hc,rel,Ao)$,其中,C 表示概念的集合,R 表示关系的集合,Hc 表示概念层次,rel 表示概念间的关系,Ao 表示本体公理.本体的描述语言很多,本文采用 W3C 的最新推荐标准 OWL 作为本体描述语言.OWL 定义了概念的语义关系如下:$\forall c_i\in C,c_j\in C$,如果 c_i 定义为 c_j 的“equivalentClass”,则称 c_i 和 c_j 语义相等,记为 $c_i\equiv c_j$,;如果 c_j 定义为 c_i 的“subClassOf”,则称 c_i 包含 c_j 语义,记为 $c_i\supseteq c_j$.由此可定义概念集合的$\equiv$和$\supseteq$语义关系:$\forall SC_i,SC_j$,如果 $(\forall c_m\in SC_i,c_n\in SC_j)\wedge(c_m\supseteq c_n\vee c_m\equiv c_n)$为真,则称 $SC_i\supseteq SC_j$;如果$(SC_i\supseteq SC_j)\wedge(SC_j\supseteq SC_i)$为真,则称 $SC_i\equiv SC_j$.

文献[16]提出可以借助服务接口组合服务的思想.基于本体,可定义服务接口语义,并将服务抽象为输入输出实体.因此,一个 Web 服务的接口信息集合可定义为 WD,服务可定义为一个二元组:$WS(DI,DO)$,其中 DI 是该服务的输入信息集合,DO 是服务的输出信息集合.服务请求可以用一个二元组表示:$WSR(DI,DO)$,其中 DI 是该请求的输入信息集合,DO 是该请求可能的输出信息集合.例如:按多个条件进行的检索可定义为一个服务请求,检索条件即为请求的输入信息集合,检索结果即为输出集合.显然,$WD\supseteq DI\wedge WD\supseteq DO$.$\forall D_i\in WD,D_i$ 可表示为一个三元组(SC,DT,R),其中:SC 是 D_i 所属的概念集合;DT 是 D_i 可能的数据实体的集合;R 是概念和数据实体间关系的集合,表达了数据实体的语义.定义信息网格中全部 Web 服务的集合为 GW,全部服务请求集合为 $GWSR$.

定义 **1**. $\forall WS_i\in GW,WS_j\in GW$,如果$(WS_i.DI.SC\supseteq WS_j.DI.SC)\wedge(WS_i.DO.SC\supseteq WS_j.DO.SC)$为真,则称 WS_i 语义包含 WS_j,记作:$WS_i\supset WS_j$.如果$(WS_i\supset WS_j)\wedge(WS_j\supset WS_i)$为真,则称 WS_i 语义等价 WS_j,记作:$WS_i\approx WS_j$.

定义 **2**. $\forall WS_i\in GW,WS_j\in GW$,如果$(WS_i.DO.DT\supseteq WS_j.DO.DT)\wedge(WS_i.DI.DT\supseteq WS_j.DI.DT)\wedge(WS_i\square WS_j)$为真,则 WS_i 包含 WS_j,记作:$WS_i\supseteq WS_j$.如果$(WS_i\supseteq WS_j)\wedge(WS_j\supseteq WS_i)$为真,则称 WS_i 等价 WS_j,记作:$WS_i\equiv WS_j$.

定义 **3**. $\forall WS_i\in GW,WS_j\in GW,WS_k\in GW$,如果$(WS_k\supseteq WS_j)\wedge(WS_k\supseteq WS_i)$为真,则称 WS_k 为 WS_i,WS_j 的一个集成,记作 $WS_i\Delta WS_j$.显然,$((WS_i\Delta WS_j)\supset WS_j)\wedge((WS_i\Delta WS_j)\supset WS_i)$为真.

命题 **1**. $\forall WS_i\in GW,WS_j\in GW$,如果$(WS_i\supset WS_j)$,则 WS_i,WS_j 存在集成.

证明:可在 WS_i 基础上构造 $WS_k(DI_k,DO_k)$,其中 $DI_k=(WS_i.DI.SC,WS_i.DI.DT\cup WS_j.DI.DT,WS_i.DI.R\cup WS_j.DI.R)$,$DO_k=(WS_i.DO.SC,WS_i.DO.DT\cup WS_j.DO.DT,WS_i.DO.R\cup WS_j.DO.R)$. 显然有 $(WS_k\supseteq WS_j)\wedge(WS_k\supseteq WS_i)$ 为真,命题得证. □

推论 **1**. 对服务序列 $WSL(WS_0,WS_1,\ldots,WS_n)$,$\forall 1\le k\le n$,如果$(WS_{k-1}\supset WS_k)$为真,则该序列存在集成.记为 Integration(*WSL*)(证明略).

定义 **4**. $\forall WS_i\in GW,WS_j\in GW$,如果$(WS_i.DO.SC\supseteq WS_j.DI.SC)$为真,则称 WS_i 和 WS_j 语义关联,记作 $WS_i\rightarrow WS_j$;如果$(WS_i.DO.SC\supseteq WS_j.DI.SC)\wedge(WS_i.DO.DT\supseteq WS_j.DI.DT)$为真,则称 WS_i 和 WS_j 服务关联,记作 $WS_i\square WS_j$.

命题 **2**. $(WS_i\rightarrow WS_j)\wedge(WS_j\rightarrow WS_i)\Rightarrow WS_i\approx WS_j$;$(WS_i\dashrightarrow WS_j)\wedge(WS_j\dashrightarrow WS_i)\Rightarrow WS_i\equiv WS_j$(证明略).

定义 **5**. $\forall WS_i\in GW,WS_j\in GW,WS_k\in GW$,如果$(WS_k.DI.SC\supseteq WS_j.DI.SC)\wedge(WS_k.DO.SC\supseteq WS_j.DO.SC)$为真,则称 WS_k 是 WS_i,WS_j 的一个语义组合,记作(WS_i+WS_j);如果$(WS_k.DI\supseteq WS_j.DI)\wedge(WS_k.DO\supseteq WS_j.DO)$为真,则称 WS_k 是 WS_i,WS_j 的一个服务组合,记作$(WS_i\pm WS_j)$.

命题 **3**. $\forall WS_i\in GW,WS_j\in GW$,如果$(WS_i\rightarrow WS_j)$,则 WS_i,WS_j 存在语义组合;$\forall WS_i\in GW,WS_j\in GW$,如果$(WS_i\dashrightarrow WS_j)$,则 WS_i,WS_j 存在服务组合.

证明:构造服务执行序列(WS_i,WS_j),执行过程中,$WS_i.DO$ 作为 $WS_j.DI$,则该序列成为一个新的服务 $WS_k(WS_i,WS_j)$. 如果 $(WS_i\rightarrow WS_j)\Rightarrow(WS_k.DI.SC\supseteq WS_j.DI.SC)\wedge(WS_k.DO.SC\supseteq WS_j.DO.SC)$;$(WS_i\dashrightarrow WS_j)\Rightarrow(WS_k.DI\supseteq WS_j.DI)\wedge(WS_k.DO\supseteq WS_j.DO)$.命题得证. □

推论 **2**. 对服务序列$(WS_0,WS_1,\ldots,WS_n)$,$\forall 1\le k\le n$,如果$(WS_{k-1}\rightarrow WS_k)$,则该序列存在语义集成.对服务序列$(WS_0,WS_1,\ldots,WS_n)$,$\forall 1\le k\le n$,如果$(WS_{k-1}\dashrightarrow WS_k)$,则该序列存在服务集成.

定义 **6**. $\forall WSR_i\in GWSR$,如果$\exists WS_i\in GW,(WS_i.DO.SC\supseteq WSR_i.DO.SC)\wedge(WS_i.DI.SC\supseteq WS_i.DI.SC)$为真,则称 WS_i 是 WSR_i 的语义实现,记作 $OReal(WSR_i)$. 如果$\exists WS_i\in GW,(WS_i.DO.SC\supseteq WSR_i.DO.SC)\wedge(WS_i.DI.SC\supseteq WS_i.DI.SC)\wedge(WS_i.DO.DT\supseteq WSR_i.DO.DT)\wedge(WS_i.DI.DT\supseteq WS_i.DI.DT)$为真,则称 WS_i 是 WSR_i 的服务实现,记作 $SReal(WSR_i)$.对服务集合 $WS=\{WS_0,WS_1,\ldots,WS_n\}$,$\forall WS_i\in WS,WS_i=OReal(WSR_i)$成立,则称 WS 是 WSR_i 的相关服务集合,记作 $Link(WSR_i)$.

在信息网格中,服务请求实现的满意度包括精确度和覆盖率两个方面.$WS_i=OReal(WSR_i)$的精确指数 E 可用公式 $E(WS_i,WSR_i)=ED(WS_i.DO,WSR_i.DO)\times EO(WS_i.DI,WSR_i.DI)$计算,其中,$EO(D_i,D_j)$表示信息集合的相似指数,$ED(D_i,D_j)=EO(D_i.SC,D_j.SC)\times Q(D_i.SC\cap D_j.SC)/Q(D_i.SC\cup D_j.SC)$,$EO(SC_i,SC_j)$表示概念集合的语义相似度;$Q(S)$表示集合 S 中的元素数量.目前,关于概念间语义相似度的计算已经有许多成果[18,19],本文借鉴这些成果给出本体的概念间语义相似度公式 $S(C_i,C_j)=a/(a+d)$.如果两个概念语义相等,则 $S=1$;如果不满足语义包含,则 $S=0$.a 是一个可调节产生,d 是一个整数.参考文献[17],本文采用以下策略:(1) $C_i\equiv C_j$ 则 $d=0,S=1$;(2) 如果 $C_i\supseteq C_j$ 或 $C_j\supseteq C_i$ 则 $d=1$.设 $SC_i=\{C_n^i\mid N\ge n\ge 0\}$,$SC_j=\{C_n^j\mid M\ge n\ge 0\}$, $EO(SC_i,SC_j)=\dfrac{1}{M\times N}\sum_{k=0}^{N}\sum_{l=0}^{M}S(C_k^i,C_l^j)$.如果 $OReal(WSR_i)$由服务序列 $WS(WS_0,WS_1,\ldots,WS_n)$构成,则

$$E(WS,WSR_i)=ED(WS_0.DI,WSR_i.DI)\times ED(WSR_i.DO,WS_n.DO)\times\prod_{i=0}^{n-1}E(WS_i,WS_{i+1})\ .$$

$Link(WSR_i)$的覆盖指数 R 定义为 $R(WS,WSR_i)=\sum_{k=0}^{N} ED(WS_k.DI,WSR_i.DI)\times Q(WS_k.DO)$.

1.2 动态集成和重构算法

动态集成和重构算法的基本思想是为信息网格中的服务请求获取最大的覆盖率和精确度.可以对 GW 集合中的服务构造集成关联图.该图可用向量矩阵表示:

$$G=\begin{pmatrix} V_{0,0} & V_{1,0} & \cdots & V_{n-1,0} & V_{n,0} \\ V_{0,1} & V_{1,1} & \cdots & V_{n-1,1} & V_{n,1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ V_{0,n-1} & V_{1,n-1} & \cdots & V_{n-1,n-1} & V_{n,n-1} \\ V_{0,n} & V_{1,n} & \cdots & V_{n-1,n} & V_{n,n} \end{pmatrix},$$

其中,$V_{ij}=(I,E)$,$E=E(WS_i,WS_j)$;$I=1$ 当且仅当$(WS_i\rightarrow WS_j)$,其他为 0.

由此,生成 WSR 服务组合即转化为加权图最短路径问题.算法伪代码如下:

WebServiceCombineList(*G*,*WSR*,min*Accurate*){ //min*Accurate* is the minimal service accurate exponent.

$$G^E=\begin{pmatrix} V_{0,0}.E & V_{1,0}.E & \cdots & V_{n-1,0}.E & V_{n,0}.E \\ V_{0,1}.E & V_{1,1}.E & \cdots & V_{n-1,1}.E & V_{n,1}.E \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ V_{0,n-1}.E & V_{1,n-1}.E & \cdots & V_{n-1,n-1}.E & V_{n,n-1}.E \\ V_{0,n}.E & V_{1,n}.E & \cdots & V_{n-1,n}.E & V_{n,n}.E \end{pmatrix};\ G^I=\begin{pmatrix} V_{0,0}.I & V_{1,0}.I & \cdots & V_{n-1,0}.I & V_{n,0}.I \\ V_{0,1}.I & V_{1,1}.I & \cdots & V_{n-1,1}.I & V_{n,1}.I \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ V_{0,n-1}.I & V_{1,n-1}.I & \cdots & V_{n-1,n-1}.I & V_{n,n-1}.I \\ V_{0,n}.I & V_{1,n}.I & \cdots & V_{n-1,n}.I & V_{n,n}.I \end{pmatrix};$$

```
For (each V_ij in G){
    WSPath=Dijkstra(G^E,WSR.DI,WSR.DO,V_ij);
        //get the optimal path from G using Dijkstra arithmetic,
    α=E(WSPath,WSR);
    if (α>minContent)WSPathSet=WSPathSet∪{WSPath};
};
For (each WSPath in WSPathSet){
    For (each WS in WSPath){
        WSTree = Prim(G^I,WS); //get the genetic tree by Prim arithmetic.
        For (each WebService path in WSTree from root to leaf){
            WSP=Integration(WSPS);
                    // WSPS is the set of WebService path in WSTree from root to leaf
            For (each WS in WSIntegrationSet)
                If (WS⊃WSP&&WSP⊃WS)WSIS=WSIS−{WS}+{WSPΔWS};
                    // WSIntegrationSet(WSIS) is the integration Web service set
        }
    }
}
Output (WSPathSet,WSIS)
}
```

按照上述算法,一个面向 WSR 的动态集成和重构的时间复杂度为 $O(k\times N^2)$,其中 k 为 $WSR.DI$ 的本体数量,N 为信息网格中 WebService 的数量.

算法保证了信息覆盖的全面性.对信息网格中的服务集合 $WS\{WS_0,WS_1,\ldots,WS_{n-1}\}$,重构的服务全集为 $IWS\{IWS_0,IWS_1,\ldots,IWS_2^n\}$.对服务请求 WSR 和服务精确度α,IWS 存在序列 $AWS(IWS_0,IWS_1,\ldots,IWS_m,IWS_{m+1},\ldots,$

IWS_2^n)满足 $E(IWS_m,WSR)\geq\alpha$,并且 $E(IWS_{m+1},WSR)<\alpha$.显然,$MWS=\{IWS_0,IWS_1,\ldots,IWS_m\}$为满足精确性要求的服务组合全集.现在需要证明算法获得的服务序列集合 $WSIS$ 与 MWS 相同.

命题 **4**. $WSIS=MWS$.

证明:$\forall IWS_i\in MWS$,按照 MWS 的定义,$E(IWS_i,WSR)\geq\alpha$,则 $IWS_i(WS_0,WS_1,\ldots,WS_n)$满足:

$$ED(WS_0.DI,WSR_i.DI)\times ED(WSR_i.DO,WS_n.DO)\times\prod_{i=0}^{n-1}E(WS_i,WS_{i+1})\geq\alpha,$$

则$\forall WS_i,WS_{i+1}\in IWS_i,WS_i\rightarrow WS_{i+1}$ 且 $E(WS_i,WS_{i+1})>0$.即 IWS_i 构成了满足 WSR 的一条集成路径.按照算法,显然有 $IWS_i\in WSIS$.$\forall WWS_i\in WSIS$, 按照算法, $ED(WS_0.DI,WSR_i.DI)\times ED(WSR_i.DO,WS_n.DO)\times\prod_{i=0}^{n-1}E(WS_i,WS_{i+1})\geq\alpha$ 且 $\forall WS_i,WS_{i+1}\in WWS_i,WS_i\rightarrow WS_{i+1}$.显然有 $WWS_i\in MWS$.命题得证. □

2 实验评估

模拟实验在 254 240 首宋诗及 9 204 个宋诗作者和 57 593 首唐诗及 6 156 个唐诗作者的相关信息、汉语大辞典 28 124 个汉字、323 399 个词以及 434 145 条解释的数据环境下进行.实验用信息网格中共有 38 个 WebService 基础服务.实验在 7 台 PⅣ2G,内存均为 512M 的 PC 服务器和局域网带宽 1G 的环境中进行,其中 6 台 PC 服务器上每个有 4~9 个 Web 服务,另一台 PC 服务器为管理服务器,运行动态集成和重构算法.数据库采用 Mysql,应用服务器为 Tomcat5.5,WebService 中间件采用 Axis1.3.实验将与全文检索(FullSearch)和数据库查询(DBSql)的覆盖率和精确度进行比较.

2.1 实验本体构造

实验系统的研究域为唐、宋代诗歌和作者,以 OWL-FULL 语言描述本体.概念包括名词性概念和谓词性概念.其中,名词概念包括:人物(people)、帝王(emperor)、唐帝王(TangEmperor)、宋帝王(SongEmperor)、诗人(poet)、僧人(monk)、时间段(time)、朝代(dynasty)、宋朝(SongDynasty)、公元纪元(AD)、年号纪元(YearName)、作品(works)、诗(poem)、小传(biography)、词条(word)、行政区划(district)、诗样式(PoemStyle)、句式(SentenceStyle)、字式(CharStyle)、平仄样式(PingZeStyle)、韵样式(YunStyle)、格律样式(GeLvStyle)、体裁样式(TiCaiStyle);谓词性概念包括:出生(birth)、死亡(dead)、创作(write)、开始(start)、结束(end)、初(head)、中(middle)、末(tail)、统治(rule).名词性概念结构如图 1 所示.

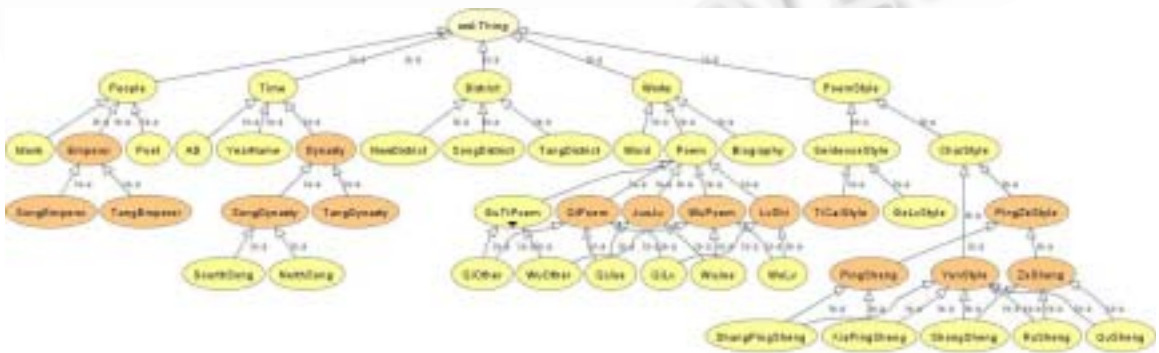


Fig.1 The relation of concept

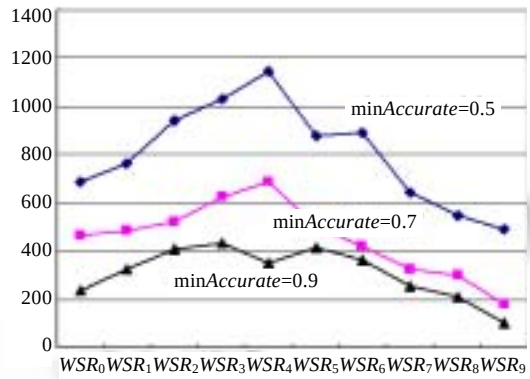
图 1 概念关系

在实验用语料中,代表同一时间语义的描述有 3 种方式,即公元、年号和朝代;代表同一地点语义的描述在不同的历史时期有所不同;代表同一诗词样式语义有不同的描述规则;代表同一人物语义可以采用不同的名称(如字、号、排行);代表同一概念语义在汉语大词典中可以有不同的词条.在实验中,唐、宋共 311 833 首诗歌作为诗(poem)概念的实例;汉语大词典的内容作为词条(word)概念的实例并用于构建同义词表.此外,本体的语义

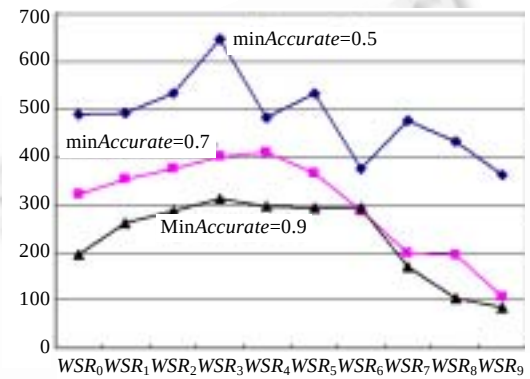
结构将对 GW 的服务构造集成关联图产生重大影响.因此,基于古典诗歌、唐宋代历史的专业知识的本体结构将直接影响服务组合的精确度和覆盖率.

2.2 WS-DIC精确度和覆盖率实验

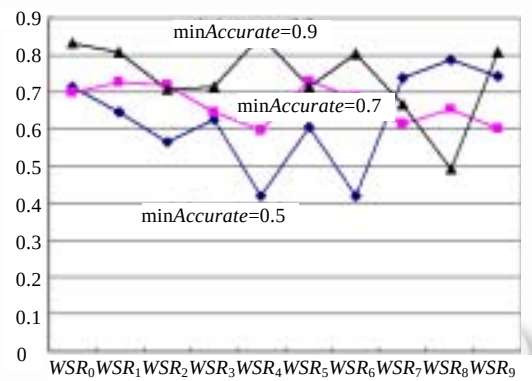
生成 10 个 *WSR* 集合的序列(WSR_0,WSR_1,…,WSR_9),组中每个 *WSR* 集合中的 *WSR* 其输入输出信息本体概念的数量和数据项均相同,本体概念数量为以 2 为差值的等差递增序列;数据项数量均为 20.每个 *WSR* 集合有 10 个 *WSR*.分别定义最低精确度为 0.5,0.7,0.9.对获得的总数据量(TQ)、相关数据量(CQ)及 RQ/CQ 的平均值进行比较,如图 2 所示.



(a) The total record quantity of WebService output (TQ)
(a) Web服务输出的总记录数量(TQ)



(b) The correlative record quantity of WebService output (CQ)
(b) Web服务输出的相关记录数量(CQ)



(c) CQ/TQ
(c) CQ/TQ

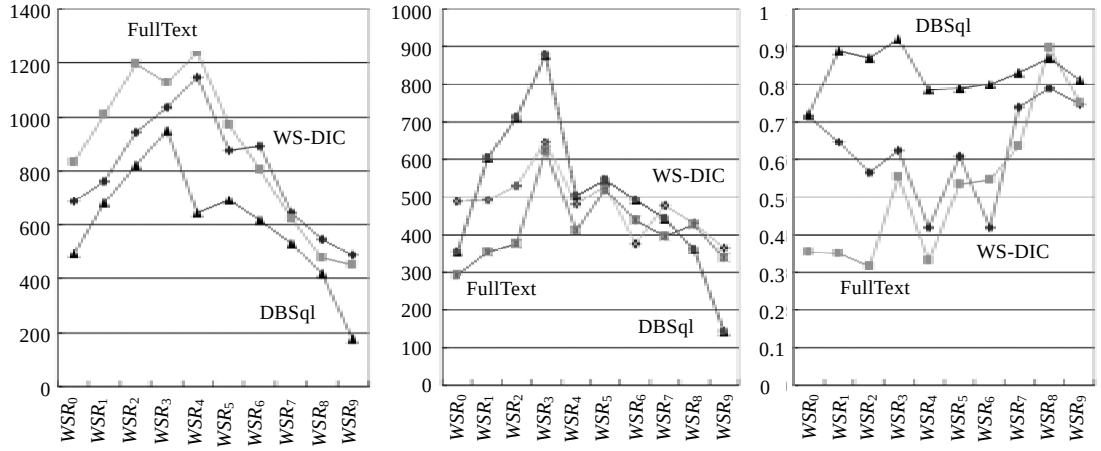
(d) The average response time (ms), min*Accurate*=0.9
(d) 平均响应时间(ms),min*Accurate*=0.9

Fig.2 The result of Experiment 1
图 2 实验 1 的结果

图 2(a)和图 2(b)表明,随着服务请求本体概念数量的递增,获得请求结果的数量先增后减.我们认为,这主要是因为当请求的本体概念较少时服务组合较多,但关联服务集成较少;但当请求的本体概念较多时,服务组合可能是唯一的,因此获取的数据集也较少.但图 2(c)表明,在获取的数据集中,有效数据集的比例基本稳定.图 2(d)的结果表明,随着请求的本体概念数量的增加,时间开销是递增的,且基本呈线性增长.

2.3 比较实验

在实验 1 的基础上,增加全文检索(FullSearch)和数据库字段查询(DBSql)两种信息搜索方式.其中,全文检索采用 Lucene 全文搜索引擎.FullSearch 和 DBSql 两种方式的 min*Accurate* 取搜索字段与请求的全部字段之比.实验结果如图 3、图 4 所示.



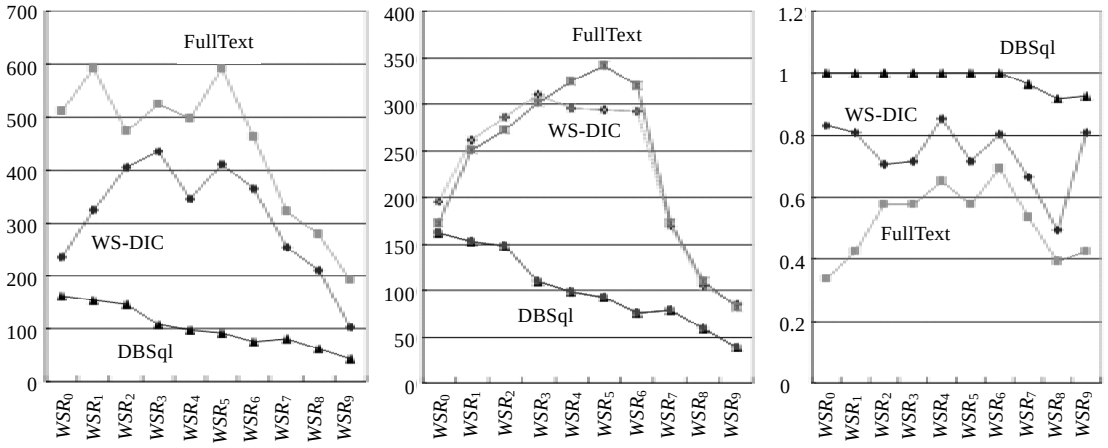
(a) The total record quantity of WebService output (TQ)
(a) Web服务输出的总记录数量(TQ)

(b) The correlative record quantity of WebService output (CQ)
(b) Web服务输出的相关记录数量(CQ)

(c) CQ/TQ
(c) CQ/TQ

Fig.3 The comparison of DBSql, FullText, WS-DIC (min*Accurate*=0.5)

图 3 DBSql,FullText,WS-DIC 在 min*Accurate*=0.5 时的比较



(a) The total record quantity of WebService output (TQ)
(a) Web服务输出的总记录数量(TQ)

(b) The correlative record quantity of WebService output (CQ)
(b) Web服务输出的相关记录数量(CQ)

(c) CQ/TQ
(c) CQ/TQ

Fig.4 The comparison of DBSql, FullText, WS-DIC (min*Accurate*=0.9)

图 4 DBSql,FullText,WS-DIC 在 min*Accurate*=0.9 时的比较

图 3(c)和图 4(c)表明:DBSql 方式获取数据的精确度最高,但其获取的有效数据最少;而 FullText 方式虽然获取了最多的结果数据,但其获取的有效数据的数量与 WS-DIC 方式相差不多.因此可以说,WS-DIC 策略在信息获取的数量和有效性方面有较大优势.

3 总结和进一步的工作

基于本体的思想解决信息网格中数据获取问题一直是研究热点.本文提出的基于本体的 Web 服务动态集成和重构策略能够最大复用已有服务,生成优化的集成和重构路径,能够有效解决信息网格中服务请求多样性和信息关联问题,同时保证了集成和重构服务的质量及效率.文中首先通过数理逻辑的理论和方法对 WS-DIC

中涉及的问题进行了描述和证明,然后借用图论的有关算法给出 WS-DIC 算法,最后在近千万项数据基础上对其进行了测试.测试结果表明,WS-DIC 与全文检索和数据库查询方式相比能够更好地满足用户的服务请求.

但是,WS-DIC 本身仍然不够完善,特别是在算法效率方面还有待改善.今后的工作将考虑如何提高 WS-DIC 算法的效率并将其在上海医学网格平台上进一步应用.上海医学网格的研究分为两个阶段:第一阶段主要是上海耳鼻喉医学网格的研究与实现,该网格部署在上海、北京两地,主要包括耳鼻喉临床辅助诊断与辅助治疗和耳鼻喉流行病学调查研究两大功能;第二阶段将进入数据量更大的乳腺癌医学网格的研究实现.当前,耳鼻喉医学网格已建设完成,第二阶段已进入设计阶段.我们计划首先在已建设完成的耳鼻喉医学网格中应用测试 WS-DIC,然后将其在乳腺癌医学网格上推广实现.

References:

[1] Xu ZW, Li XL, You GM. Architecture study of the VEGA information grid. Journal of Computer Research and Development, 2002,39(12):948−951 (in Chinese with English abstract).

[2] Li XL, Xu ZW, Liu XW, Yang N. Community-Based model and access control for information grid. In: Proc. of the IEEE/WIC Int'l Conf. on Web Intelligence 2003 (WI 2003). Halifax: IEEE Computer Society, 2003. 462−465.

[3] Foster I, Kesselman C. The Globus project: A status report. In: Proc. of the 7th Heterogeneous Computing Workshop (HCW'98). Orlando: IEEE Computer Society, 1998. 4−18.

[4] Basney J, Livny M. Managing network resources in condor. In: Proc. of the 9th Int'l Symp. High-Performance Distributed Computing. Washington: IEEE Computer Society, 2000. 298−299.

[5] Li W, Xu ZW, Bu GY, Zha L. An effective resource locating algorithm in grid environments. Chinese Journal of Computers, 2003, 26(11):1546−1549 (in Chinese with English abstract).

[6] Rajasekar A, Wan M, Moore R, Kremenek G, Guptil T. Data grids, collections, and grid bricks. In: Proc. of the 20th IEEE/11th NASA Goddard Conf.on Mass Storage Systems and Technologies (MSST 2003). Washington: IEEE Computer Society, 2003. 2−9.

[7] Casati F, Ilnicki S, Jin LJ, Krishnamoorthy V, Shan MC. eFlow: A platform for developing and managing composite e-services. In: Proc. of the Academia/Industry Working Conf. on Research Challenges. Buffalo: IEEE Computer Society, 2000. 341−348.

[8] Meng J, Krithivasan R, Su SYW, Helal S. Flexible inter-enterprise workflow management using e-services. In: Proc. of the 4th IEEE Int'l Workshop on Advanced Issues of E-Commerce and Web-Based Information Systems (WECWIS 2002). Newport Beach: IEEE Computer Society, 2002. 43−50.

[9] Meng J, Su SYW, Lam H, Helal A. Achieving dynamic inter-organizational workflow management by integrating business processes, events and rules. In: Proc. of the 35th Annual Hawaii Int'l Conf. on System Sciences (HICSS 2002). Big Island: IEEE Computer Society, 2002. 10.

[10] Mcguinness DL, Fikes R, Hendler J, Stein LA. DAML+OIL: An ontology language for the semantic Web. Intelligent Systems, 2002,17(9-10):72−80.

[11] McIlraith SA, Son TC, Zeng HL. Semantic Web services. Intelligent Systems, 2001,16(3-4):46−53.

[12] Wang HB, Zhang YQ, Sunderraman R. Soft semantic Web services Agent. In: Proc. of the IEEE Annual Meeting of Fuzzy Information (NAFIPS 2004), Vol 1. Banff: IEEE Computer Society, 2004. 126−129.

[13] Nau D, Au TC, Ilghami O, Kuter U, Wu D, Yaman F, Munoz-Avila H, Murdock JW. Applications of SHOP and SHOP2. Intelligent Systems, 2005,20(3-4):34−41.

[14] Liang QA, Chung JY, Miller S. Towards semantic service request of Web service composition. In: Proc. of the IEEE Int'l Conf. on e-Business Engineering (ICEBE 2005). 2005. 705−712. http://doi.ieeecomputersociety.org/10.1109/ICEBE.2005.121

[15] Bechhofer S, Harmelen F, Hendler J, Horrocks I, Deborah L, Guinness M, Peter F, Schneider P, Stein LA. OWL Web ontology language reference. 2006. http://www.w3.org/TR/owl-ref/

[16] Majithia S, Walker DW, Gray WA. Automated Web service composition using semantic Web technologies. In: Proc. of the Int'l Conf. on Autonomic Computing. IEEE Computer Society, 2004. 306−307. http://csdl.computer.org/comp/proceedings/icac/2004/2114/00/21140306.pdf

[17] Li M, Wang DZ, Du XY, Wang S. Dymamic compositon of Web services based on domain ontology. Chinese Journal of Computers, 2005,28(4):644 −650 (in Chinese with English abstract).

[18] Rosso P, Masulli F, Buscaldi D. Word sense disambiguation combining conceptual distance, frequency and gloss. In: Proc. of the Int'l Conf. on Natural Language Processing and Knowledge Engineering. 2003. 120−125. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=1275880

[19] Navigli R, Velardi P. Structural semantic interconnections: A knowledge-based approach to word sense disambiguation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005,27(7):1075−1086.

附中文参考文献:

[1] 徐志伟,李晓林,游赣梅.织女星信息网格的体系结构研究.计算机研究与发展,2002,39(12):948−951.

[5] 李伟,徐志伟,卜冠英,查礼.网格环境下一种有效的资源查找方法.计算机学报,2003,26(11):1546−1549.

[17] 李曼,王大治,杜小勇,王珊.基于领域本体的 Web 服务动态组合.计算机学报,2005,28(4):644 −650.

陈磊(1976－),男,河南信阳人,博士生,工程师,主要研究领域为网格计算技术,高性能计算.

韩颖(1978－),女,博士生,主要研究领域为现当代文学,古典诗歌.

李三立(1935－),男,教授,博士生导师,中国工程院院士,CCF 高级会员,主要研究领域为网格计算技术,高性能计算技术.