

基于本体概念集合相似度的语义 Web 服务匹配

杨佳^{1,2}, 张金广^{1,2}, 杨龙^{1,2}, 江萍^{1,2}, 魏晓莉^{1,2}

(1. 合肥工业大学管理学院, 安徽合肥 230009;

2. 教育部过程优化与智能决策重点实验室, 安徽合肥 230009)

摘要:面向应用领域的 Web 服务发现技术的目标是帮助用户找到能够为其构建应用系统提供功能支持的 Web 服务, 从而降低应用系统的开发成本和提高开发效率, 而在服务发现技术中服务匹配程度的高低是一个关键。为了更有效准确地获得服务之间的匹配程度, 文章从 Web 服务的输入与输出参数的语义描述出发, 把语义 Web 服务的匹配问题转化成本体库中概念向量的相似度计算问题, 通过在考虑概念结构上的路径长度、深度和密度影响因素上计算两个概念的相似度的基础上, 构建了概念集合的相似度算法。实验表明该算法较别的匹配方法在查全率、特别是查准率上得到了提高。

关键词: Web 服务匹配; 语义相似度; Web 服务描述

中图分类号: TP31

文献标识码: A

文章编号: 1673-629X(2012)08-0056-04

Semantic Web Services Matching Based on Ontology Concept Set Similarity

YANG Jia^{1,2}, ZHANG Jin-guang^{1,2}, YANG Long^{1,2}, JIANG Ping^{1,2}, WEI Xiao-li^{1,2}

(1. School of Management, Hefei University of Technology, Hefei 230009, China;

2. Key Laboratory of Process Optimization and Intelligent Decision-making of Ministry of Education, Hefei 230009, China)

Abstract: The goal of application-oriented Web service discovery technology is to enable customers to find out the suitable Web services so as to support functionally the building of application systems, thus reducing cost and improving efficiency, but the degree of service matching is a key in the service discovery technology. In order to get an effective and accurate matching degree of discriminating services, based on the semantic description of Web Service's input and output, the problem of matching semantic Web services is transformed to the computation of semantic similarity between concepts vector in domain ontology. Through considering path length, depth and density factors of the conceptual structure to calculate the similarity of two concepts, thus construct the set of concept similarity algorithm. Experiment results show the enhancement to accuracy and recall.

Key words: Web services matching; semantic similarity; Web services description

0 引言

随着 Web 服务数量的逐渐增加, 如何快速准确地发现目标服务已成为一个迫切需要解决的问题, 而 Web 服务匹配是服务发现质量的关键。传统的服务匹配技术是在采用语法级别的描述语言 WSDL (Web services description language)^[1] 的基础上, 通过关键词匹配来实现, 没有考虑到资源内部的语义信息以及概念间的关系问题, 使得查全率与查准率较低, 从而降低了匹配结果的质量。

目前, 解决这个问题的途径就是通过融合传统 Web 服务与语义技术, 将 Web 服务语义信息加入到 Web 服务描述中去, 如 OWL-S (Web ontology language service)^[2]。在 OWL-S 语言中, 一个 Web 服务的功能是通过输入参数、输出参数、服务执行前提条件、服务执行结果 (IOPE) 来描述^[3], 服务输入、服务输出参数中用到的词汇概念称为术语, 这些术语关系通过所属的本体概念表述。在 Web 服务使用 OWL-S 描述的前提下, 服务匹配其实是基于本体概念的匹配^[4], 而概念之间的匹配过程是通过某种方式来选取最接近的概念, 这可以通过语义相似度来度量它们的匹配程度。

文中主要利用本体论分别对用户请求服务与发布服务进行语义描述, 通过单个本体概念的相似度计算得到服务的输入与输出匹配度, 最终得出匹配结果。

收稿日期: 2011-12-26; 修回日期: 2012-03-29

基金项目: 国家“863”云制造主题项目 (2011AA040501); 国家自然科学基金 (70871033); 安徽省教育自然科学基金重点项目 (KJ2011A006)

作者简介: 杨佳 (1987-), 女, 硕士研究生, 研究方向为人工智能、机器学习。

1 相关概念

1.1 本体(Ontology)

在哲学领域中,本体的定义是“对世界上客观存在物的系统描述”,是对客观存在的一个系统的解释和说明。在人工智能领域中,它的定义是“给出构成相关领域词汇的基本术语和关系,以及利用这些术语和关系构成的规定这些词汇外延的规则的定义”,是用一个特定的词汇对实际存在的事物的描绘。

目前比较公认的本体定义是:本体是对某个领域共享的概念化的一个详细描述^[5]。它定义了用在特定领域的术语,可以用来描述特定领域的知识。它可以形象地表示为一个网络图的结构。例如,一个用于描述交通工具的本体可以用图1直观地表现出来。

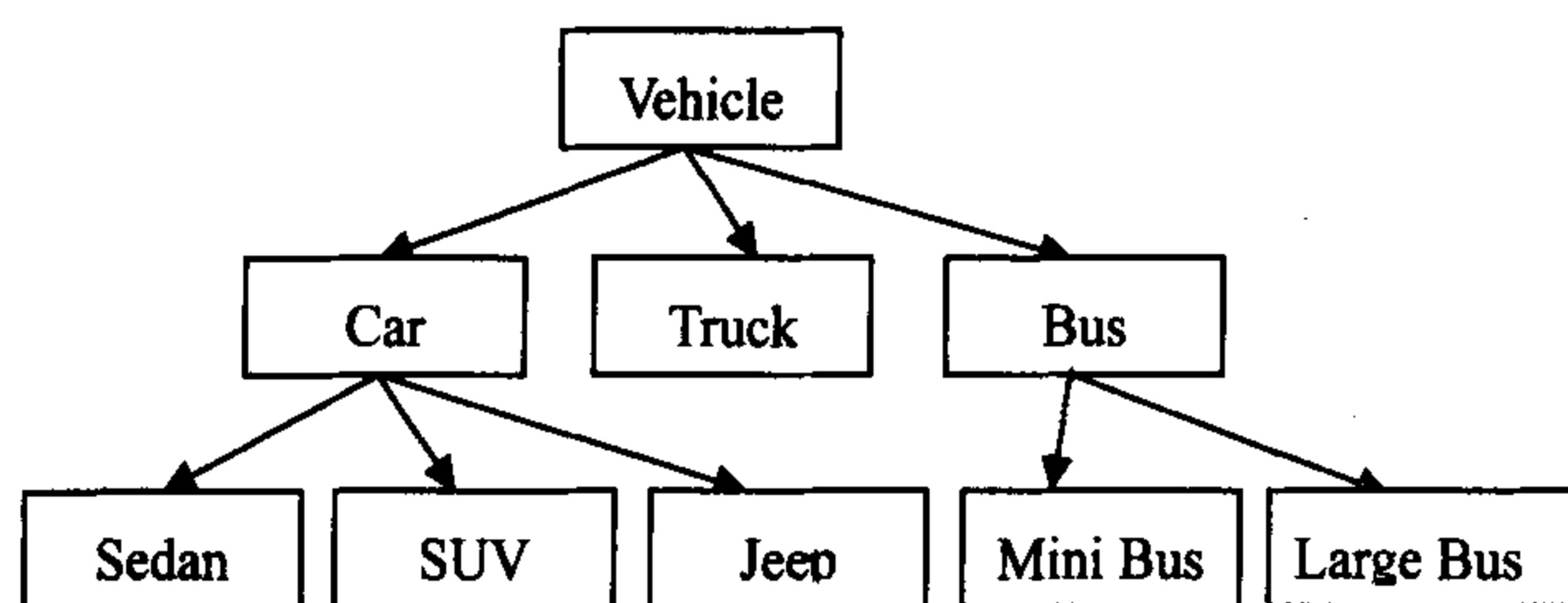


图1 交通工具的本体

1.2 Web 服务描述

Web 服务描述可分为服务的功能属性与非功能属性的描述^[6],目前对其描述有 OWL-S、WSDL-S、SWSL 等多种,其中发展最成熟且应用最广泛的是 OWL-S。OWL-S 是基于本体的描述 Web 服务属性和功能的规范。在 OWL-S 中,一个服务主要由 Service Profile、Service Model 和 Service Grounding 这三种来描述。Service Profile 描述服务主要做什么;Service Model 描述服务是怎么做;Service Grounding 描述怎么访问服务。

对服务的功能属性描述即 Web 服务的 Input、Output、Precondition、Effect(IPOE)进行语义描述是先用本体表达领域内的语义信息,再利用本体中的概念对 Web 服务的功能信息进行描述。对服务的非功能属性的描述一般用自然语言方法进行语义描述。

文中是采用 OWL-S 语言对 Web 服务描述,并且只考虑对服务的功能属性的语义匹配。特别地,由于大部分 Web 服务描述文档中缺少对 Precondition、Effect 的描述,因而在文中只考虑 Web 服务的 Input、Output 参数的语义匹配,主要表现为服务需求方与服务发布方的输入与输出的匹配。在文中,Web 服务语义描述模型定义为对服务功能信息中的输入和输出参数的语义上进行的描述,它可以形象地表示为 $WS = \langle I, O \rangle$ 。其中: $I = \langle I_1, I_2, \dots, I_n \rangle$ 是一概念向量,表示 Web 服务 WS 的 n 个输入参数的语义描述; $O = \langle O_1, O_2, \dots, O_m \rangle$ 也是一概念向量,表示 Web 服务 WS

的 m 个输出参数的语义描述; I_1, I_2, \dots, I_n 表示各个输入参数在领域本体库中相对应的概念; O_1, O_2, \dots, O_m 表示各个输出参数在领域本体库中相对应的概念。

2 基于概念相似度的 Web 服务匹配算法

2.1 相关服务匹配算法分析

Paolucci 等人提出的 OWL-S/UDDI 匹配算法^[7]是一个经典算法,它将本体概念之间的匹配程度分为 4 个不同的等级(精确匹配、包含匹配、嵌入匹配以及失败匹配)。下面以发布服务与请求服务的输出为例:

设请求服务的输出概念为 outR,发布服务的输出概念为 outA。

若 outR = outA 或者 outR 是 outA 的直接子类,则属于精确匹配;

若 outR 包含 outA,则属于包含匹配;

若 outA 包含 outR,则属于嵌入匹配;

若上述条件都不成立,则属于失败匹配。

这一算法具有局限性:它只是 4 个离散的匹配程度,不能给出服务之间的精确匹配,且不是所有没有包含关系的服务都为失败匹配。文献[8]通过定义一个分段函数计算本体中的两个概念相似度,并与 OWL-S/UDDI 算法的匹配结果进行比较,但概念的相似度匹配程度仍然不高。文献[9]考虑了概念之间继承关系、二元关系以及概念深度对相似度的影响,但其算法只能应用在由单个本体概念构成的服务接口。当服务接口是由多个概念组成的集合则需要根据概念集合的相似度进行匹配。

2.2 改进的 Web 服务匹配算法

文中通过计算请求服务与发布服务的输入集合与输出集合的相似度来确定请求服务语义描述与发布服务语义描述的匹配程度,即将其问题转化为计算同一领域本体中概念向量 I 和 I' 以及 O 和 O' 的相似度。两个概念向量集合的相似度是在计算两个概念的语义相似度的基础上用余弦算法^[10]来计算。

具体的算法步骤主要分为以下部分:

(1) 计算两个概念之间的相似度。

两个概念的语义相似度是在考虑概念间路径长度、概念深度和概念密度因素的基础上计算。设 l 是概念 c_1 与概念 c_2 之间的最短路径长度, h 是 c_1 与 c_2 的深度最大的共同祖先的深度, d 是 c_1 与 c_2 在本体中的密度,则两个概念 c_1, c_2 的语义相似度 $\text{sim}(c_1, c_2)$ 为:

$$\text{sim}(c_1, c_2) = \begin{cases} 1, & \text{if } c_1 = c_2 \\ e^{-al} \times \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}}, & \text{if } c_2 \notin \text{Siblings}(c_1) \\ e^{-al-d} \times \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}}, & \text{if } c_2 \in \text{Siblings}(c_1) \end{cases}$$

其中, α 与 β 是平滑因子, 取值在 $0, 1$ 之间。

①在第一种情况, 当两个概念完全相同时, 它们相似度为 1 , 意味最高相似程度。

②在第二种情况, 当两个概念为非兄弟结点时, 考虑了两个概念的最短路径长度以及它们共同祖先的深度这两个因素的影响, 它们的相似度随着最短路径长度的增加而递减, 且在最短路径长相等下, 它们的相似度随着共同祖先的深度增加而递增。

③在第三种情况, 当两个概念为兄弟结点时, 在考虑第二种情况的影响因素下, 还考虑了结点密度的影响, 即两个概念在同深度下路径长度相等下, 结点密度越高的区域比密度低的区域的相似程度高。

(2) 计算请求与 Web 服务输入与输出匹配度。

在同一个领域本体 O 中, 根据用户输入 I 与输出 O 可以建立用户请求输入向量与用户请求输出向量, 根据已有的服务 S 的输入参数 I' 与输出参数 O' 可以建立发布服务输入向量与发布服务输出向量。

令与 I 相关的概念集合为 $RI = \{r \in O \mid \exists i \in I: \text{sim}(r, i) > 0\}$, 与 O 相关的概念集合为 $RO = \{r \in O \mid \exists o \in I: \text{sim}(r, o) > 0\}$ 。

则用户请求输入向量为:

$$QI = \langle QI_1, QI_2, \dots, QI_n \rangle$$

其中 $QI_x = \max(\text{sim}(QI_x, i))$ 且 $1 \leq x \leq n, i \in I$ 。

用户请求输出向量为:

$$QO = \langle QO_1, QO_2, \dots, QO_n \rangle$$

其中 $QO_x = \max(\text{sim}(QO_x, o))$ 且 $1 \leq x \leq n, o \in O$ 。

发布服务输入向量为:

$$AI = \langle AI_1, AI_2, \dots, AI_n \rangle$$

其中 $AI_x = \max(\text{sim}(AI_x, i'))$ 且 $1 \leq x \leq n, i' \in I'$ 。

发布服务输出向量为:

$$AO = \langle AO_1, AO_2, \dots, AO_n \rangle$$

其中 $AO_x = \max(\text{sim}(AO_x, o'))$ 且 $1 \leq x \leq n, o' \in O'$ 。

通过分别计算用户请求输入、输出向量和发布服务输入、输出向量的相似度来得到请求与 Web 服务的输入、输出匹配度, 两向量的相似度一般是通过向量之间的夹角余弦值来计算, 即:

$$\begin{aligned} \text{sim}_{\text{input}} &= \text{sim}(QI, AI) = \frac{QI \cdot AI}{|QI| \times |AI|} \\ &= \frac{\sum_{x=1}^n QI_x \times AI_x}{\sqrt{\sum_{x=1}^n QI_x^2} \times \sqrt{\sum_{x=1}^n AI_x^2}} \end{aligned}$$

$$\text{sim}_{\text{output}} = \text{sim}(QO, AO) = \frac{QO \cdot AO}{|QO| \times |AO|}$$

$$= \frac{\sum_{x=1}^n QO_x \times AO_x}{\sqrt{\sum_{x=1}^n QO_x^2} \times \sqrt{\sum_{x=1}^n AO_x^2}}$$

$\text{sim}_{\text{input}}$ 与 $\text{sim}_{\text{output}}$ 的取值在 $(0, 1)$ 之间, 它的值越高意味着它们的匹配程度越高。

(3) 最终匹配结果。

通过将请求与 Web 服务输入与输出匹配度分别乘以相关系数进行求和来得到综合的匹配度作为最终的匹配结果。综合匹配度 sim 计算为:

$$\text{sim} = w_1 \times \text{sim}_{\text{input}} + w_2 \times \text{sim}_{\text{output}}$$

其中 w_1 与 w_2 分别是请求与 Web 服务输入与输出匹配度的权重系数, 且它们之和为 1 (一般可以设 $w_1 < w_2$, 这样可以加大输出相似性在整个计算中的权重)。

若综合匹配度值大于 0 , 则将 Web 服务加入到结果集中, 最后按照其值由大到小进行排序将结果集返回给用户。

3 实验结果分析

文中实验中的 Web 服务、用户请求以及本体都是从开源的 OWLS-TC 数据集^[11]中获得的。OWLS-TC 数据集中包含了 8 个领域有 1000 多条服务, 从中随机地抽取了 4 条服务, 并采用 5 个匹配要求, 主要是从查准率与查全率这两方面来测试的, 并将文中方法的实验结果与 OWL-S/UDDI 匹配算法的结果进行了对比。在一个查询 Q 中, $S = \{S_1, S_2, \dots, S_m\}$ 是查询结果集, S_q 是用户所需服务集, $C = \{C_1, C_2, \dots, C_m\}$ 是查询条件集, $R = \{R_1, R_2, \dots, R_n\}$ 是在条件 C 中的结果集, 则查全率 QCR 与查准率 QAR 的计算式^[12]为:

$$\text{QCR} = \frac{|S_q|}{|S_1 \cup S_2 \cup \dots \cup S_m|} \times 100\%$$

$$\text{QAR} = \frac{|S_q|}{|R|} \times 100\%$$

图 2 给出了两种算法在 5 个查找请求下平均查准率与查全率之间的关系。其中横轴表示查全率 QCR, 纵轴表示查准率 QAR。

从图 2 中可以看出, 文中提出的 Web 服务匹配算法的查准率明显高于 OWL-S/UDDI 算法。当查全率为 0.1 左右的时候, 它们的查准率都接近 1.0 , 都能找到符合精确匹配的服务。但是, 之后随着查全率的提高, OWL-S/UDDI 算法的查准率迅速降低而文中的算法的准确率仍然较高, 这是由于 OWL-S/UDDI 算法只划分四个匹配等级使得只要是与请求语义上稍微相关

就会被认为是符合要求的匹配,而文中的算法不仅考虑概念因素的影响还结合向量上余弦计算使得查准率提高。

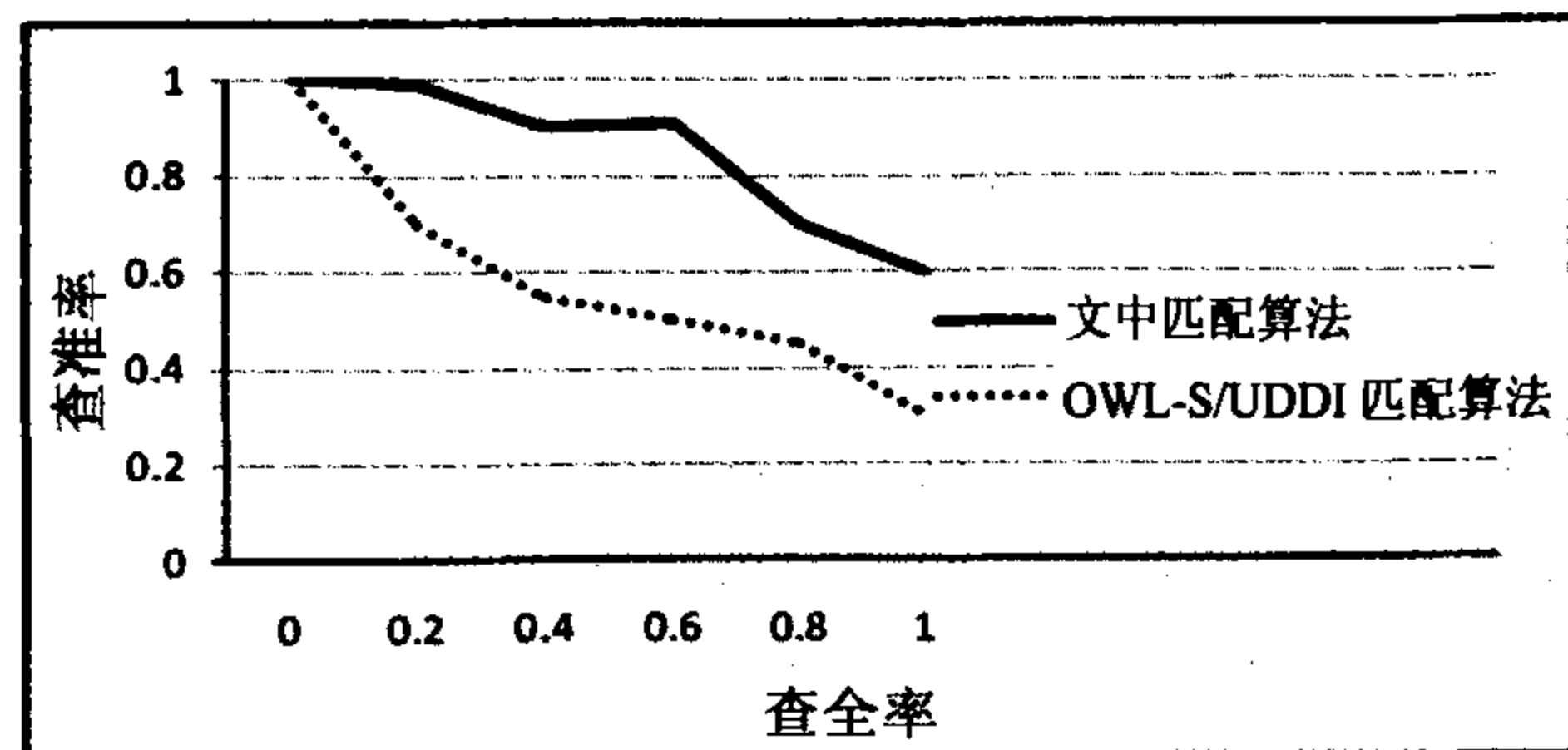


图2 QCR-QAR 曲线

4 结束语

文中提出了一种新的语义 Web 服务匹配算法,通过在本体中计算两个概念之间的相似度基础上计算概念向量的相似度得到 Web 服务输入与输出匹配度,最后将各分量匹配度乘以权值求和得到综合匹配度,将其排序后的结果集返回给用户。根据实验的结果显示了该算法显著提高了服务匹配准确率。

参考文献:

- [1] Christensen E, Curbera F, Meredith G, et al. Web services description language (WSDL) 1.1 [EB/OL]. 2001. <http://www.w3.org/TR/2001/NOTE-wsdl-20010315>.
- [2] Martin D, Burstein M, Hobbs J, et al. OWL-S: semantic mark-

up for web services [EB/OL]. 2004. <http://www.w3.org/Submission/OWL-S/>.

- [3] 于敏,曹宝香. 基于 QoS 的语义 Web 服务匹配策略[J]. 计算机技术与发展, 2010, 20(9): 78-82.
- [4] 黄志成,李华. 改进的语义 Web 服务匹配算法设计与实现[J]. 计算机工程, 2009, 35(20): 88-90.
- [5] 聂规划,左秀然,陈冬林. 本体映射中一种改进的概念相似度计算方法[J]. 计算机应用, 2008, 28(6): 1563-1565.
- [6] 张献,李舟军,李梦君. 一种基于语义 web 服务匹配的策略和实现[J]. 计算机科学, 2007, 34(5): 99-103.
- [7] Paolucci M, Kawamura T, Payne T, et al. Semantic matching of web services capabilities [C]//Lecture Notes in Computer Science. [s.l.]: [s.n.], 2002: 333-347.
- [8] 彭晖,史忠植,邱莉榕,等. 基于本体概念相似度的语义 Web 服务匹配算法[J]. 计算机工程, 2008, 34(15): 51-53.
- [9] 杨惠荣,刘珊珊,尹宝才,等. 基于语义距离的 Web 服务匹配算法[J]. 北京工业大学学报, 2011, 37(4): 591-595.
- [10] Bonino D, Corno F. Self-similarity Metric for Index Pruning in Conceptual Vector Space Models [C]//19th International Conference on Database and Expert Systems Application. [s.l.]: [s.n.], 2008: 225-229.
- [11] Klusch M, Fries B, Khalid M, et al. Owls-mx: hybrid owl-s service match making [EB/OL]. 2005. <http://www.aaai.org/Papers/Symposia/Fall/2005/FS-05-01/FS05-01-011.pdf>.
- [12] Yang Zhi, Chen Junliang, Wu Budan. A New Ontology-based Service Matching Algorithm [C]//2010 6th World Congress on Services. [s.l.]: [s.n.], 2010: 170-171.

(上接第 55 页)

4 结束语

文中阐述了 Windows 环境下的 PCI-E 驱动设计和实现,该驱动采用 WDF 驱动模型开发,该模型提供了比以往的 WDM 更高层次抽象的高度灵活、可扩展、可诊断的驱动程序框架,封装了驱动中的某些共同行为,并将驱动程序与操作系统之间进行了分离,因此简化了驱动程序设计人员的开发难度,缩短了开发周期。笔者对该驱动程序在自研的某芯片验证板卡上进行了功能和性能测试,经测试,各接口功能正确,可以稳定、可靠的工作;DMA 读速率为 460MB/s, DMA 写速率为 570MB/s,驱动接口功能及性能满足该芯片的 PCI-E 接口验证的需求。

参考文献:

- [1] 马鸣锦. PCI、PCI-X、PCI Express 的原理及体系结构[M]. 北京:清华大学出版社, 2007.
- [2] PCI Express 2.0 Base Specification [EB/OL]. [2006-09-1]. <http://www.pcisig.com>.

- [3] PCI local bus specification v3.0 [EB/OL]. [2010-03-10]. <http://www.pcisig.com>.
- [4] 张帆,史彩成. Windows 驱动开发技术详解[M]. 北京:电子工业出版社, 2008.
- [5] 李正军. WDF 设备驱动程序的设计与实现[J]. 计算机技术与发展, 2007, 17(5): 228-230.
- [6] Orwick P. Developing Drivers with the Microsoft Windows Driver Foundation [M]. US: Microsoft Press, 2007.
- [7] 程海全,胡君. PCI 设备电源管理驱动程序开发[J]. 现代电子技术, 2010(14): 196-201.
- [8] 廖寅龙,田泽. FC 网络通信中 PCIe 的接口的设计与实现[J]. 航空计算技术, 2010, 40(4): 127-130.
- [9] 武安河. Windows 设备驱动 WDF 开发[M]. 北京:电子工业出版社, 2009.
- [10] Microsoft Windows driver kit 7600 [EB/OL]. [2009-10-01]. <http://www.microsoft.com>.
- [11] Oney W. Programming The Microsoft Windows Driver Model (WDM) [M]. US: Microsoft Press, 1999.
- [12] 黎邵秀. PCI-E 图像采集系统的 WDF 驱动程序设计[J]. 科学技术与工程, 2011, 11(16): 3824-3832.

作者: 张海亮, 袁进华
作者单位: 四川大学计算机学院, 四川成都610065
刊名: 计算机技术与发展
英文名: Computer Technology and Development
年, 卷(期): 2012(8)

参考文献(12条)

1. Christensen E;Curbera F;Meredith C Web services description language(WSDL) 1.1 2001
2. Martin D;Barstein M;Hobbs J OWL-S:semantic markup for web services 2004
3. 于敏;曹宝香 基于QoS的语义Web服务匹配策略[期刊论文]-计算机技术与发展 2010(09)
4. 黄志成;李华 改进的语义Web服务匹配算法设计与实现[期刊论文]-计算机工程 2009(20)
5. 葛规划;左秀然;陈冬林 本体映射中一种改进的概念相似度计算方法[期刊论文]-计算机应用 2008(06)
6. 张敏;李舟军;李梦君 一种基于语义web服务匹配的策略和实现 2007(05)
7. Paolucci M;Kawamura T;Payne T Semantic matching of web services capabilities 2002
8. 彭晖;史忠植;邵蔚榕 基于本体概念相似度的语义Web服务匹配算法[期刊论文]-计算机工程 2008(15)
9. 杨惠荣;刘珊珊;尹宝才 基于语义距离的Web服务匹配算法[期刊论文]-北京工业大学报 2011(04)
10. Bonino D;Corno F Self-similarity Metric for Index Pruning in Conceptual Vector Space Models 2008
11. Klusch M;Fries B;Khalid M Owls-mx:hybrid owl-s service match making 2005
12. Yang Zhi;Chen Junliang;Wu Budan A New Ontology-based Service Matching Algorithm 2010

本文链接: http://d.g.wanfangdata.com.cn/Periodical_wjz201208014.aspx