

# 基于 SPARQL-DL 的语义 Web 服务查询

王海<sup>1</sup>, 高岭<sup>1</sup>, 范琳<sup>2</sup>, 李增智<sup>3</sup>

(1. 西北大学信息科学与技术学院, 陕西西安 710069; 2. 西安邮电学院计算机学院, 陕西西安 710061;  
3. 西安交通大学计算机系统结构与网络研究所, 陕西西安 710049)

**摘要:** 语义 Web 服务发现是当前 Web 服务研究领域的热点, 其核心研究内容是服务描述及相应的发现方法. 服务描述分为请求描述与发布描述, 发布描述具有描述完整, 信息丰富的特征, 而服务请求描述仅关注服务的部分特征, 通常不构成一个完整的服务描述. 现有方法使用相同机制进行请求描述与发布描述, 以比对请求与发布中对应部分的同一性或相似性作为匹配依据. 构建假想的完整服务描述作为服务请求既不合理也难以实施, 从而限制了方法的实用性. 本文提出以语义 Web 查询语言 SPARQL-DL 作为服务请求描述语言, 以 OWL-S 作为服务发布描述语言, 将服务发现问题转化为知识库的查询, 从而进行有效的服务发现. 通过实验, 证实了该方法具有实用性强, 简单可靠的特点.

**关键词:** 语义 Web 服务; 服务查询; SPARQL-DL

**中图分类号:** TP311      **文献标识码:** A      **文章编号:** 0372-2112 (2011) 3A-052-05

## Query for Semantic Web Service Based on SPARQL-DL

WANG Hai<sup>1</sup>, GAO Ling<sup>1</sup>, FAN Lin<sup>2</sup>, LI Zeng-Zhi<sup>3</sup>

(1. School of Information Science and Technology, Northwest University, Xi'an, Shanxi 710069, China;  
2. Computer Science and Technology Department, Xi'an Institute of Posts and Telecommunications, Xi'an, Shanxi 710061, China;  
3. Institute of Computer System Structure and Networks School, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China)

**Abstract:** The Semantic Web service discovery is a hot spot in the current web service research field. Its core research content is service descriptions and the corresponding discovery methods. Service description can be divided into the service request description and advertisement description. Service advertisement descriptions are usually complete, descriptive, information-rich descriptions; while service request description is concerned only with some of the characteristics of services, usually does not constitute a complete description of the service. Current methods of service discovery using the same mechanism to describe both request and advertisement. Compare and identify the identity or similarity of the corresponding part between the two as matching basis. Build a complete fictive service description as service request limit the practicability of the method. In this paper, we propose to use semantic web query language SPARQL-DL as the service request description language to retrieve OWL-S form published services. Treat the service discovery issue as query against the knowledge base. Through the experiment, confirmed that the method is practical, simple, reliable and easy-to-use.

**Key words:** semantic web services; service query; SPARQL-DL

## 1 引言

Web 服务架构把分布式的异质网络资源整合成松散耦合的、与平台及语言无关的无缝集成系统, 是解决企业异质信息系统集成以及满足企业动态需求变化的有效途径. 为了实现这一目标, 首要解决的问题是 Web 服务的描述和发现机制. Web 服务发现即根据服务使用者 (service requester) 提供的服务请求, 找到满足请求的服务. 作为实现服务共享、复用的关键支撑技术, 受到学术界和工业界的广泛关注.

服务发现的核心研究问题集中在服务描述及相应的发现方法. 服务描述从使用角度可以分为请求描述和发布描述; 相应发现方法通常分为两个步骤, 即服务匹配和结果评级 (或称结果排序). 已有服务发现研究工作, 主要集中在服务匹配方法, 相应的提出了各种服务描述, 但并未将服务请求描述与服务发布描述的描述机制区分开来. 本质上这两者具有天然的不对称性, 服务发布描述通常是一个完整的服务描述, 全面而细致; 与此相反, 服务请求描述往往只描述服务请求者关注的部分, 通常不构成一个完整的服务描述. 考虑到这样的信

息不对称性,现有的服务发现方法假定服务请求与发布采用相同描述,以比较对应的部分同一性或相似性作为匹配依据,即不合理,也难实际应用。

本文根据目前语义 Web 的研究进展,提出以语义 Web 查询语言 SPARQL-DL<sup>[1]</sup> 构建服务请求描述,利用 OWL-S<sup>[1~3]</sup> 作为服务发布描述,进行有效的语义 Web 服务发现。

## 2 相关工作

### 2.1 相关研究

针对以 UDDI 为代表的,关键字匹配为依据的服务发现技术查全率与查准率不高的问题,学术界已有不少相关研究.文献[4]提出以描述逻辑 TBox 包含性检测为依据,对服务请求描述与服务发布描述的输入输出进行相容性检测,得到精确,包含,插拔和否定四种匹配类型.文献[5]以描述逻辑的 ABox 一致性检测为工具,对服务请求描述与服务发布描述做一致性检测,得到精确,包含,插拔,相交和否定五种匹配类型.描述逻辑开放世界语义推理要求所有知识必须显示声明,当服务功能较复杂时,会造成知识库爆炸.文献[6]中,作者通过一个关键语义相似度计算函数,根据服务不同属性综合计算服务之间相似度,提高了查全率与查准率.文献[7]对描述逻辑添加 AAX 认知模态词 K,采用局部封闭世界语义.在描述较复杂服务时,克服了知识库爆炸问题.文献[8]将 WSDL 描述中的操作(Operation)建模为无序标签数,通过计算满足一定约束的编辑距离来度量操作之间的相似度,作为匹配依据,实质上是基于相似度计算的方法.

从上述研究基础可以看出,各种服务发现方法都考虑了有效使用服务描述的语义信息,从而提高查全率与查准率.文献[4,5,7]的方法,以假想的完整服务描述作为服务请求描述,与发布服务描述中对应部分做比对以得到匹配结果.没有考虑到请求服务描述与发布服务描述之间的不对称性,特别是请求服务描述的特点,虽然查全率与查准率高,但是由于需要给出完整的假想请求服务描述的所有细节,方法难以实际应用.文献[8]虽然考虑到了服务请求描述与服务发布描述之间的不对称性,但没有充分认识到服务请求的本质就是一组条件约束,而非全面的服务特性刻画,其匹配是以 WSDL 描述中操作(Operation)的树形结构编辑距离计算为基础,实质上是一种相似度计算,其查全率与查准率难以保障。

### 2.2 OWL DL 及其语义

OWL-DL 尽管是在 RDF 基础上构建的,它却与 RDF 有着非常不同的语义.形式上,一个 OWL-DL 的本体词汇由一个七元组组成,即  $V_O = (V_{cls}, V_{op}, V_{dp}, V_{ap}, V_{ind},$

$V_D, V_{lit})$ ,其中  $V_{cls}$  表示概念名的 URI 集合,  $V_{op}$  表示对象属性的 URI 集合,  $V_{dp}$  表示数据类型属性的 URI 集合,  $V_{ap}$  表示标注属性的 URI 集合,  $V_{ind}$  表示个体的 URI 集合,  $V_D$  表示数据类型名称的 URI 集合,  $V_{lit}$  表示 RDF 文字的集合. OWL-DL 提供了一组算子用以构造复杂概念. OWL-DL 概念可以归纳的如下定义:

$$C \leftarrow A \mid \text{not}(C) \mid \text{and}(C_1, \dots, C_2) \mid \text{or}(C_1, \dots, C_2) \\ \mid \{a\} \text{some}(p, C) \mid \text{all}(p, C) \text{min}(n, q) \mid \text{max}(n, q) \\ \mid \text{some}(t, D) \mid \text{all}(t, D) \mid \text{min}(n, t) \mid \text{max}(n, t)$$

其中,  $A \in V_{cls}$ ,  $C_{(i)} \in Sc$ ,  $a \in V_{ind}$ ,  $p \in V_{op}$ ,  $t \in V_{dp}$ ,  $D \in V_D$ ,  $n$  表示一个非负整数。

OWL-DL 的语义则采用一个塔斯基的解释结构来给出.一个解释表示为  $I = (D^I, \cdot^I)$ ,其中  $D^I$  表示论域,有互不相交的目标域  $D'_0$  和数据域  $D''_0$  组成.  $I$  表示解释函数,它给本体中的每个实体一个确定的含义.对于每一个概念  $C \in V_{cls}$ ,  $I$  将其映射为一个  $D'_0$  的子集  $C^I_0$ ; 每一个对象属性  $p \in V_{op}$ , 映射为一个二元关系  $P^I \subseteq D'_0 \times D'_0$ ; 每一个数据类型属性  $t \in V_{dp}$ , 映射为  $t^I \subseteq D'_0 \times D''_0$ ; 每一个标注属性  $p_a \in V_{ap}$ , 映射为  $p^I \subseteq D^I \times D^I$ ; 每一个个体  $a \in V_{ind}$  映射为  $a^I \in D'_0$ ; 每一个文字  $l \in V_{lit}$  映射为  $l^I \in D''_0$ . 其它复杂概念的解释可以由以上扩展而得,具体可以参考<sup>[1]</sup>.

### 2.3 OWL-S

OWL-S<sup>[2,3]</sup> 是一种以 OWL 定义的 Web 服务本体.它提供了一套核心标记语言,以一种明确的、计算机可理解的方式描述 Web 服务的属性和功能. ServiceProfile, 如图 1 描述服务能做什么,类似于黄页信息,包括服务的描述性信息,以及服务的输入、输出、前置条件和效果,即 IOPE. ServiceModel 用以刻画服务如何工作,而 ServiceGrounding 则给出服务如何使用的细节。

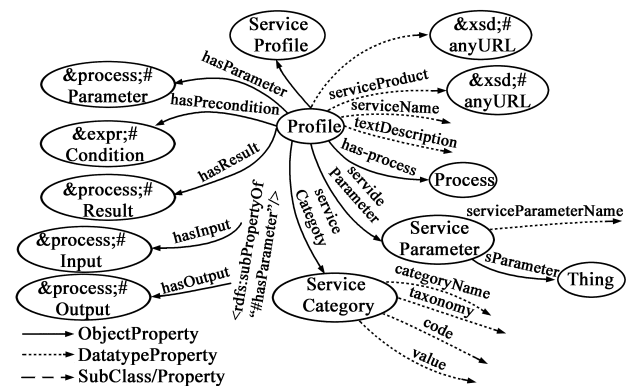


图1 OWL-S Profile

### 2.4 SPARQL-DL

令  $V_O = (V_{cls}, V_{op}, V_{dp}, V_{ap}, V_{ind}, V_D, V_{lit})$  表示一个 OWL-DL 的本体词汇集. 并且令  $V_{bnode}$  和  $V_{var}$  表示 bnode 标识符集合和变量集合. 一个 SPARQL-DL 原子查询  $q$  形式如图 2。

$q^{-1}Type(a, C)$   
 $PropertyValue(a, p, v)$   
 $SameAs(a, b)$   
 $DifferentFrom(a, b)$   
 $Equivalent(C_1, C_2)$   
 $SubClassOf(C_1, C_2)$   
 $DisjointWith(C_1, C_2)$   
 $ComplementOf(C_1, C_2)$   
 $EquivalentProperty(p_1, p_2)$   
 $SubPropertyOf(p_1, p_2)$   
 $InverseOf(p_1, p_2)$   
 $ObjectProperty(p)$   
 $Functional(p)$   
 $InverseFunctional(p)$   
 $Transitive(p)$   
 $Symmetric(p)$   
 $Annotation(s, p_a, o)$

图2 SPARQL-DL 原子查询形式

其中,  $a, b \in V_{uri} \cup V_{bnode} \cup V_{var}$ ,  $v \in V_{uri} \cup V_{lit} \cup V_{bnode} \cup V_{var}$ ,  $p, q \in V_{uri} \cup V_{var}$ ,  $C, D \in S_c \cup V_{var}$ ,  $s \in V_{uri}$ ,  $p_a \in V_{ap}$ ,  $o \in V_{uri} \cup V_{lit}$ . 一个 SPARQL-DL 的查询  $Q$  一个有穷的原子查询集合, 查询被解释为这些原子查询的合取. SPARQL-DL 查询, 分为五类: (1) 标准的 TBox 查询; (2) 标准的 ABox 查询; (3) 带有自由变元的 ABox 查询; (4) ABox 与 TBox 混合查询; (5) ABox 与 RBox 混合查询.

对于描述逻辑推理机难以处理的项值约束 (term value constraint), SPARQL-DL 中使用外置的过滤器可以方便的解决. 例如给定的本体中, 某个概念  $C$  有属性 HasName 和 HasPrice, 查询中要限制 Price 的值, 带过滤器的查询如图 2.

### 3 一个网上书店的例子

该例子来自 Amason.com, 其服务以语义 Web 服务的形式对外提供, 在此不失一般性, 及论文篇幅限制, 我们选取四个典型的经过简化的服务作为测试用例, 如表 1 所示, 服务功能从其输入输出观察不言自明: (服务实际以 OWL-S 形式发布):

表 1 网上书店服务

服务名	输入	输出
SearchByTitle	book: Title	book: Book
SearchByAuthor	book: Author	book: Book
GetBookPrice	book: Book	book: Price
GetBookAuthor	book: Book	book: Author

使用 SPARQL-DL 对服务描述进行查询:

(1) 对于想对书目进行搜索的用户, 可能想得到书目搜索的服务, 那么要查询这样的服务, 其查询语句如下:

```

select? s
{
  ? s rdf: Type service: Service
  ? s service: is PresentedBy? p
  ? p profile: hasInput? i
  ? i process: parameterType? it
  ? it rdfs: subclassOf? book: Book
}

```

查询结果如下: ? s GetBookPrice GetBookAuthor

(2) 对于获取给定书目信息相关的服务, 其查询语句如下所示:

```

select? s
{
  ? s rdf: Type service: Service
  ? s service: isPresentedBy? p
  ? p profile: hasOutput? o
  ? o process: parameterType? ot
  ? ot rdfs: subclassOf? book: Book
}

```

查询结果如下: ? s SearchByTitle SearchByAuthor

(3) 对于获取给定书目的价格的服务, 其查询如下:

```

select? s
{
  ? s rdf: Typeservice: Service
  ? s service: isPresentedBy? p
  ? p profile: hasInput? i
  ? p profile: hasOutput? o
  ? i process: parameterType? it
  ? o process: parameterType? ot
  ? it rdfs: subclassOf? book: Book
  ? ot rdfs: subclassOf? book: Price
}

```

查询结果如下: ? s GetBookPrice

### 4 结果分析

上述三类查询恰恰反映了针对服务查询的三种基本查询类型, 即

(1) 给定输出信息的服务查询, 此种查询的目标是发现那些可能得到输出信息的有关服务, 如上例的查询 1, 其结果就是得到输出为给定书目的服务集合;

(2) 给定输入信息的服务查询, 此种查询的目标是发现那些与输入信息有关的服务. 例如上述查询 2, 发现那些与书目有关的服务;

(3) 同时给定输入输出信息的服务查询, 这类查询的目的性很明确, 就是要得到特定的服务. 如上例的查

询 3,其结果就是得到了按照给定书目查询书目价格的服务。

以 SPARQL-DL 语义 Web 查询语言来进行 OWL-S 的语义服务的请求描述,较以往的方法具有明显的优势:

(1)将服务发布描述与服务请求描述的描述机制明确区分开来.因为服务发布描述一般需要完整的描述服务的各个方面,尽量全面;而服务请求描述并不需要完整的描述服务的各个方面,只需要给出请求约束部分即可.以往的发现方法中,都是以构造完整的请求服务描述作为请求约束,这即不合理,也难以应用。

(2)SPARQL-DL 查询语言具有丰富的表达能力,其不仅仅能够查询 OWL 表达的各种显式信息,还查询到 OWL 表达的隐式信息,并且内置了数值运算,逻辑比较操作以及简单的类 SQL 的统计功能,可以处理服务的各种非功能属性约束。

## 5 实验及分析

为了对本文提出的基于 SPARQL-DL 的服务发现方法进行比较评价,使用 OWL-S 服务发现测试集 OWL-S TC V2<sup>[9]</sup>.该测试集包含符合 OWL-S 1.1 推荐标准的服务描述 570 多个,覆盖 7 个应用领域.分别是教育、医疗保险、食品、旅行、通信、经济和军事.OWL-S TC V2 同时提供了 28 个测试查询集(即服务请求),每个查询与 10~20 个不等的服务相关.作者对这 28 个测试查询以 SPARQL-DL 形式重新构建,并进行了实验.为了对比分析,同时将这 28 个测试查询按关键字进行了提取,并按照文献[6]的做法进行了以关键字相似度为依据的发现.并且,也以文献[4]中的描述逻辑包含性检测为基础的服务发现方法进行了测试。

由于 OWL-S 本质上是 OWL 语言的一个应用,而 SPARQL-DL 与 OWL 的描述是相容的,在 SPARQL-DL 可以直接引用 OWL 中的描述.因此,服务发现系统类似一个数据库查询系统结构,如图 3 所示。

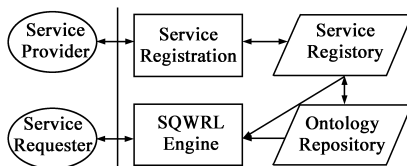


图3 服务发现测试系统结构

从图 4 及图 5 可以看出,在查全率指标上,以 SPARQL-DL 构建服务请求,比基于包含性检测的方法小幅改善,但是相对关键字的方法,有大幅度提升,查准率的情况与此类似.主要原因在于,SPARQL-DL 形式的服务请求,可以涵盖包含性检测,并同时支持 QoS 约束,考虑到使用 SPARQL-DL 过滤器.因而相对包含性检

测类方法,有小幅改善.关键字方法由于仅仅考虑词汇相似度,以关键字的自然语言语义为基础,在处理二义性及多义性上存在困难,因此其发现结果不够理想。

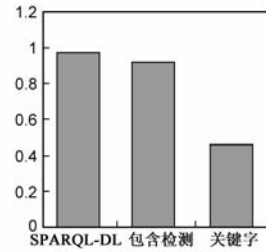


图4 查全率对比图

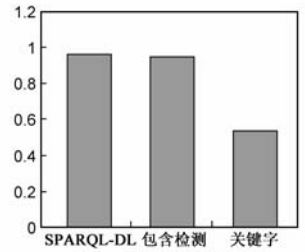


图5 查准率对比图

## 6 结束语

语义 Web 服务发现,实际是对语义 Web 服务描述发现.服务描述分为发布描述和请求描述,这两者的关系并不对称.本文以语义 Web 查询语言 SPARQL-DL 作为服务请求描述语言,对 OWL-S 形式发布的服务进行发现.最后以 OWL-S TC V2 标准语义服务测试集合进行了实验,结果证明该方法查全率与查准率有所提升,并具有实用性强,简单可靠的特点。

## 参考文献

- [1] Sirin E, Parsia B, SPARQL-DL: SPARQL Query for OWL-DL, in Golbreich C, Kalyanpur A, Parsia B, eds. Proceedings of the OWLED 2007, <http://ceur-ws.org/Vol-258/paper14.pdf>.
- [2] David Martin, Mark Burstein, et al, OWL-S: Semantic Markup for Web Services [EB/OL], available at <http://www.daml.org/services/owl-s/1.1/overview/>, 2004.
- [3] OWL-S Coalition, OWL-S [EB/OL], available at <http://www.daml.org/services/owl-s/>, 2005.
- [4] Paolucci M, Kawamura T. Semantic matching of Web services capabilities [A]. In Proceedings of the first International Semantic Web Conference (ISWC2002) [C]. Sardinia, Italy: Springer-Verlag GmbH, 2002. 333 - 347.
- [5] L Li, I Horrocks. A software framework for matchmaking based on semantic web technology [C]. In WWW'03: Proceedings of the 12th international conference on World Wide Web [C]. New York, NY, USA: 2003. ACM.
- [6] 丘田,李鹏飞,林品.一个基于概念语义近似度的 Web 服务匹配算法 [J]. 电子学报, 2009, 37(2): 429 - 432.  
Qiu Tian, Li Pengfei, Lin Pin. A Web service matching algorithm based on semantic similarity of concepts, Acta Electronica Sinica, 2009, 37(2): 429 - 432. (in Chinese)
- [7] Grimm S, B Motik, et al. (2006). Matching Semantic Service Descriptions with Local Closed-World Reasoning. The Semantic Web: Research and Applications: 575 - 589.
- [8] 何玲娟,刘连臣,吴澄.一种改进的基于 WSDL 描述的操作

作相似性度量方法, 计算机学报, 2008, 31(8): 1331 - 1339.

He Lingjuan, Liu Lianchen, Wu Cheng. A modified operation similarity measure method based on WSDL description, Chinese Journal of Computers, 2008, 31(8): 1331 - 1339.

- [9] Sem Web Central. Owls-TC: An OWL-S service retrieval test collection[EB/OL]. <http://projects.semwebcentral.org/projects/owls-tcV2>, 2008.

## 作者简介



王海 男, 博士, 1977 年生于陕西西安, 西北大学信息科学与技术学院讲师. 研究方向为服务计算、语义 Web、智能信息系统.

E-mail: hwang@u.edu.cn



高岭 男, 博士, 西北大学信息科学与技术学院教授, 博士生导师, 中国计算机学会高级会员, 网络与数据通信专业委员会委员, IEEE Member, 陕西省高校电化教育学会副理事长, 陕西省信息化领导小组专家顾问委员会委员, 陕西省信息与安全保密专家顾问组成员. 研究方向为计算机网络、服务计算.

范琳 女, 1982 年生于湖北钟祥, 西安邮电学院计算机学院讲师, 研究方向为嵌入式 Internet、服务计算.

李增智 男, 1938 年生于陕西渭南, 西安交通大学计算机科学技术系教授, 博士生导师, 研究方向为计算机网络管理、服务计算.