

一种本体增强的云服务发现方法

杨健,朱斌,刘宁,李冠宇

(大连海事大学 信息科学与技术学院,辽宁 大连 116026)

摘要:现有的云服务发现方法是基于关键字搜索,且手动地查询用户需求的云服务,其查询结果中会包含不相关的信息,容易遗漏相关的服务,增加用户工作量且不能满足用户的一些高级个性化查询要求。为了解决现有云服务发现方法中的这些问题,提出了一种语义物联网中本体增强的云服务发现方法。该方法利用云本体的概念匹配技术,借助理性推理机对云服务的描述信息和用户需求的描述信息进行相似性推理,增强了用户需求与云服务之间的匹配精度。通过对本体增强的云服务发现框架与算法的描述以及实验后的性能分析表明,该方法可以为服务请求者提供更加准确的服务,提高了服务发现的成功率并高效地发现满足用户需求的云服务。

关键词:云计算;云服务;云本体;本体增强;云服务发现

中图分类号:TP18

文献标识码:A

文章编号:1673-629X(2016)08-0001-06

doi:10.3969/j.issn.1673-629X.2016.08.001

A Method of Ontology-enhanced Cloud Services Discovery

YANG Jian, ZHU Bin, LIU Ning, LI Guan-yu

(College of Information Science & Technology, Dalian Maritime University, Dalian 116026, China)

Abstract: Existing cloud service discovery method is based on keyword searching, and the cloud services need to be checked by the user manually. Query results in this method contain irrelevant information, which leads to miss the relevant services. At the same time, it cannot satisfy the user high-grade demand and increase in the workload of the user. In order to solve these problems, a method of ontology-enhanced cloud services discovery is proposed, which takes advantage of concepts matching technology of cloud ontology. With the help of inference engine, the similarity between cloud ontology descriptions and user demands is calculated. Description of framework, algorithm and experimental performance analysis show that the new method can provide more accurate services for service requesters. Moreover, it can efficiently discovery some cloud services which meet the demands of users.

Key words: cloud computing; cloud service; cloud ontology; ontology-enhanced; cloud service discovery

0 引言

随着云计算的广泛应用,云服务资源的种类随之不断增加,云服务提供者以互联网为基础媒介向云服务使用者提供云资源及服务,但是云服务使用者却很难有效地使用^[1]。这是因为到目前为止,还没有一个明确的方法可以有效地从多种多样的云服务中发现最符合用户需求的云服务。

现有的云服务发现方法是基于关键字搜索,以传统的搜索引擎作为查询机制。用户查找到云服务后,还需要手动查看关于云服务的描述信息^[2],既而找到最满足需求的云服务。该方法不仅增加了用户的工作量,而且很容易遗留与需求相关的云服务。因此,用户

需要一种能够更加快速、高效地查找出满足需求的新方法。

针对上述问题,文中提出了语义物联网中本体增强的云服务发现方法。该方法在云服务的发现过程中,创建云本体,将本体、概念之间的关系应用于云服务的发现过程中,实现云计算领域内知识的推理与重用,使计算机更好地解析概念和属性所表达的真实含义。

云服务的描述信息与用户需求的描述信息借助理性推理机进行相似性推理与匹配后,可以提高两者之间的相似程度,有效地为用户提供满足个性化需求的云服务。

收稿日期:2015-11-19

修回日期:2016-03-03

网络出版时间:2016-08-01

基金项目:国家自然科学基金资助项目(61371090)

作者简介:杨健(1990-),女,硕士研究生,CCF会员,研究方向为智能信息处理;朱斌,副教授,硕士,研究方向为智能信息处理;刘宁,教授,硕士,研究方向为智能信息处理;李冠宇,教授,博士,研究方向为智能信息处理。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160801.0904.022.html>

1 基本概念

1.1 云计算

云计算(Cloud Computing)^[3]是一种新型的计算模式,以互联网作为工具提供动态的,且容易扩展的虚拟化资源,同时可以对相关服务进行增加、使用 and 交付等操作。云计算^[4]对云外的用户提供不同层次与级别的多种服务,并且可按需动态配置服务。云计算特有的以“一切资源即服务”为中心的服务方式,即一切资源都是以服务的形式提供给用户选择和使用,使得云计算成为继语义网之后的又一流行词。

云计算的服务模式^[5]主要有:

(1)软件即服务(Software as a Service, SaaS):服务使用者通过 Internet 访问服务提供者开发的,且以服务模式提供的软件功能,使用者不需要关心软件所在主机的底层基础结构具体是什么形式。

(2)平台即服务(Platform as a Service, PaaS):指将开发软件所使用的平台作为一种服务资源,以 SaaS 的模式交给服务使用者进行使用。

(3)基础设施即服务(Infrastructure as a Service, IaaS):服务使用者以 Internet 作为基础从完整的计算机基础设施中获得服务。

1.2 云服务

云计算以互联网为媒介提供动态的,易于扩展的虚拟化服务资源。将网络、互联网比喻成“云”,抽象地表示互联网和底层基础设施。云服务指以网络为基础,以按照需求的方式来获取自身所需要的服务,特点是按照需求进行分配且容易进行扩展^[6-7]。它意味着可以通过互联网将计算和存储能力以商品的形式进行流通。

简单地讲,云服务将企业发展所需要的软硬件和资料都存放在网络上,无论什么时间、地点,只要使用设备都能互相连接,就能实现数据的储存、获取、运算等目的。

到目前为止,使用较多的云服务有公有云(Public Cloud)和私有云(Private Cloud)两种,见表 1。

表 1 公有云与私有云的区别

云服务	区别
公有云	云服务中最基本的服务,多个服务使用者可共享一个服务提供商的系统资源,使用者无需架设和构建任何设备以及配备管理职员,便能享有专业的 IT 服务。公有云对于一般创业者、中小企业来说,是一个降低成本的好办法。目前常用的 Gmail/Hotmail/网上相册都是常用的公有云
私有云	运作形式与公有云类似。企业需要自己架设数据中心、网络和构建存储设备,并且需要拥有专业的顾问团队。因此构建公司自己的私有云是一项重大投资

1.3 本体与云本体

1.3.1 本体

本体,又被称为本体论^[8],被知识工程学者当成一种能够在语义和知识层面上描述信息系统的概念建模的工具。

1993 年,Gruber 给出了一个最为流行的定义:本体是概念模型的明确的规范说明。这个定义包含的四层含义如表 2 所示。

表 2 本体定义的四层含义

含义	描述
概念化 (Conceptualization)	通过抽象出客观世界中某些现象的相关概念而得到的概念模型
明确化(Explicit)	所有概念及其相关约束都有明确的定义
形式化(Formal)	本体是可以被计算机处理的
共享(Share)	本体反映了相关概念中公认的概念集合

概括地说,本体即为概念以及概念之间的关系^[9]。

1.3.2 云本体

云计算作为一种新型的计算模式,允许用户通过网络临时使用各种计算机基础设施。云计算的提供商将这些基础设施抽象成一个或多个抽象层,每层提供一种服务。软件应用程序、编程平台、数据存储、基础设施和硬件都被抽象成一种服务,这些服务被看作是云计算服务的核心,但是其内部关系十分模糊复杂导致内部互操作非常困难。而且云计算领域也缺少对内部依赖关系的理解,阻碍了云计算领域研究的进一步发展。因此,Youseff^[10-11]等尝试着为“云”建立一个详细的本体,并提出了介绍云计算领域的知识、组件及其之间关系的新概念。

文中所提出的云本体,是为了查找基础设施即服务中提供的服务而创建的云计算领域本体。建立云本体可以帮助用户更好地理解云计算模型中各个组件之间的内部关系,以及如何利用其他的云组件重新组合现有的系统来满足某种新的需求^[12]。

图 1 为云本体中云概念之间的层次化关系。云本体包括软件应用层、编程平台层、硬件层以及云系统层。其中,云系统层为主要层次,包括平台即服务层、基础设施即服务层、数据即服务层、软件即服务层以及通信即服务层。

2 本体增强的云服务发现

2.1 本体增强的云服务发现的总体框架

云服务提供者通过注册中心注册发布自己所提供的云服务,并在注册服务的同时提供关于云服务的描述信息。另一方面,云服务需求者通过 Web 界面提交自己的请求。当发现平台收到该请求信息后,首先对

其进行预处理,处理成系统可以识别的需求格式,并将处理结果发送到推理及匹配模块。当推理匹配模块收到经过预处理的需求信息后,便与发布在注册中心的云服务描述信息进行推理与匹配。在匹配过程中参考云本体,借助语义推理机对两者进行相似性推理,获得概念之间的相似程度,增强服务匹配精度。经过需求信息预处理,需求信息与服务信息的匹配,服务过滤与评分,选出满足用户请求的服务,将结果通过 Web 界面返回给用户参考。

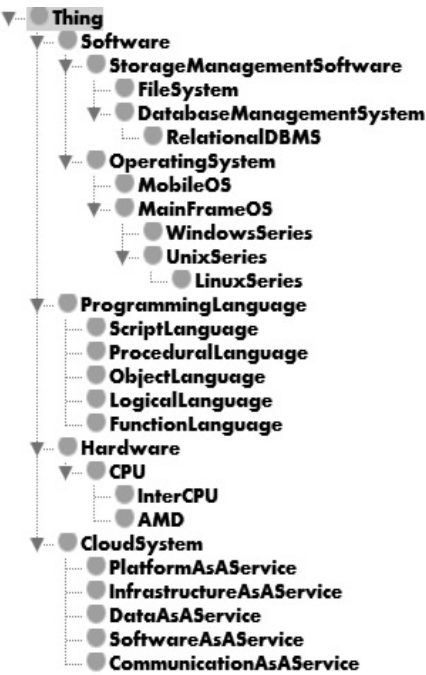


图 1 云本体中概念之间的层次化关系图

由此,图 2 提出了本体增强的云服务发现框架这一概念性模型。该框架的核心功能主要包括 Web 界面、注册中心、信息预处理、服务匹配以及服务的评估与过滤。

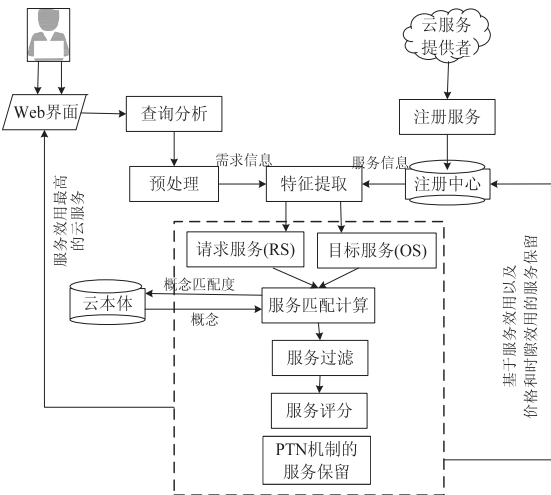


图 2 本体增强的云服务发现框架

(1) Web 界面。

是用户和系统交互的接口。云服务提供者以标准

的属性规格注册自己的云服务,云服务需求者通过 Web 界面以标准的属性规格向系统提交自己的服务请求。例如,用户需要选择自己所需的操作系统、CPU 等一系列信息。系统收到用户的请求后,请求将被发送到预处理模块进行处理。当云服务发现过程完成时,若该服务与用户的请求一致,则通过 Web 接口返回给用户,供用户选择和使用。

(2) 注册中心。

注册中心包含已注册服务的提供者及其所提供的服务,同时也包含了云服务提供者和需求者制定的每一种服务方案的价格效用、时隙效用和平均效用的基本信息。这些基本信息用于计算 PTN(Price and Time-slot Negotiation)效用,精确决定云服务的保留。

(3) 预处理。

收到 Web 界面发送过来的用户需求信息后,预处理阶段将服务请求信息进行分词处理,并将服务请求信息转化成系统可以识别的向量格式,称之为请求服务(RS),发送到服务匹配模块,与云服务提供者在注册中心注册发布的服务即目标服务(OS)进行匹配。在预处理的过程中,需要进行词法分析、停用词删除等操作。主要目的是提取用户查询请求中有用的服务信息,使得经过该模块处理后的服务请求信息更能精确地表达用户的需求。

预处理阶段的执行过程主要包括四个步骤:

- Web 界面收到用户提交的需求信息后将查询请求信息发送至推理机。
- 推理机分别对每个查询请求信息查询云本体,判定查询请求信息是否存在于云本体中,把不存在于云本体中的查询条件或参数剔除掉。
- 将收到的查询请求信息进行词法分析、分词等操作,进行标准化,以达到本体增强的云服务发现系统可以识别的统一表现形式的目的。
- 规范化、可以识别的服务请求描述发送给语义服务匹配引擎。

(4) 服务匹配。

语义推理机根据服务信息和已发布注册的服务信息,运用本体中的概念,概念间所包含的各种关系,以及推理机的推理功能来计算二者的匹配程度,最后将匹配结果返回给服务匹配引擎。文中设计了针对三种不同情况的云服务匹配方法,计算两个服务之间的相似程度。

(5) PTN 原理的服务评分与过滤。

对满足用户需求的云服务进行评分与过滤,服务效用是决定服务评分的主要因素。PTN 原理方便需求者和提供者之间进行服务协商。云服务发现过程参考查阅所构建的云本体,以便找到最符合用户个性化需

求的服务。PTN 原理提供给云服务的需求者多种依据价格和时间的服务方案。在云服务的发现过程中, PTN 原理和服务过滤一起被执行, 用户可以看到不同服务的价格和时隙的效用, 选出满足自己偏好的最优服务。

2.2 云服务发现的算法描述以及推理方法

云服务的发现过程可以描述如下: 用户提交服务查询请求信息, 云服务发现系统首先预处理用户的需求信息, 处理成系统可以识别的格式, 并将其发送到服务匹配模块。在接收到请求后, 服务匹配模块计算请求服务信息和已注册的云服务信息之间的相似程度。经过连续不断地处理、匹配, 提高了过滤效果, 找到满足用户需求的服务。

云本体提供的元数据信息, 描述了数据的语义, 包括云概念和云个人及其之间的关系。云服务发现系统接收到用户提交的参数信息后, 参考查询构建云本体来实现服务的推理和匹配。

对于服务之间的匹配, 利用构建的云本体, 有以下三种情况^[13-15]: 相似性推理、等价性推理、数据类型属性相似性推理。

(1) 相似性推理。

通过统计两个概念节点可达的共同节点来计算。概念的相似性代表了概念之间的共性。概念相似性的计算如式(1)所示:

$$\text{Sim}_{\text{comp}}(x, y) = \rho \frac{|\alpha(x) \cap \alpha(y)|}{|\alpha(x)|} + (1 - \rho) \frac{|\alpha(x) \cap \alpha(y)|}{|\alpha(y)|} \quad (1)$$

其中, ρ 为影响度, 取值范围为 $[0, 1]$; $\alpha(x)$ 是节点 x 向上可达的节点集合; $\alpha(x) \cap \alpha(y)$ 是节点 x, y 向上可达的节点交集, 是概念 x 和概念 y 共性的象征。

$\text{Sim}_{\text{comp}}(x, y) = 0$ 表示概念 x 与 y 不相似, $\text{Sim}_{\text{comp}}(x, y) = 1$ 则表示概念 x 和 y 相等。但是, 相似性推理不能比较两个兄弟节点之间的相似性。

(2) 等价性推理。

等价性推理是基于标签值来决定兄弟节点概念之间的相似性。例如计算两个软件版本之间的兼容性, 用两个兄弟节点代表不同的软件版本, 经过相似性推理会得到很高的相似度, 但是依据兄弟节点代表时间先后顺序的标签值来计算, 就会得到不同的相似度。两个兄弟节点概念之间的相似性如式(2)所示:

$$\text{Sim}_{\text{equ}}(x, y) = \text{Sim}(x, y) + \frac{0.8 |c_1 - c_2|}{10} \quad (2)$$

其中, c_1, c_2 分别为概念 x 和概念 y 的标签值; $\text{Sim}(x, y)$ 为推理时得到的相似度值。

(3) 数据类型属性相似性推理。

数据类型的相似性推理用来决定两个数据类型概念之间的相似性。相同领域的数据类型相似性计算如式(3)所示:

$$\text{Sim}_{\text{data}}(x, y) = 1 - \left| \frac{x - y}{\text{Max}_c - \text{Min}_c} \right| \quad (3)$$

其中, x 为第一个数据类型的属性值; y 为第二个数据类型的属性值; c 为概念名; Max_c 和 Min_c 为数字属性的取值范围。

若提供服务与云服务之间的匹配满足上述匹配方法之一, 且得到的匹配度大于规定的阈值, 则进行下一个概念的比较, 直到完成所有的概念和属性的比较。若服务的所有概念和属性都满足用户的需求, 则将服务信息通过 Web 接口返回给用户, 由用户来选择最佳的服务。如果不满足上述情况, 则中止比较。

云服务发现算法的伪代码如下所示:

```

对所有经过预处理后得到的请求服务结果集合:  $\{R(1), R(2), \dots, R(N)\}$ 
1. while(注册中心目标服务数量>0)
    计算目标服务集合  $\{s(1), s(2), \dots, s(N)\}$  中  $s(1)$  和请求服务集合  $\{R(1), R(2), \dots, R(N)\}$  中  $R(1)$  的相似性:
        Step1: 相似性推理
        Step2: 等价性推理
        Step3: 数据类型推理
2. 如果两个概念经过 Step1 得到相同的相似度, 则它们是兄弟节点
    那么进行 Step2
3. 如果两个概念是数据类型的值
    那么进行 Step3
4. 否则进行 Step1
5. 经过步骤 2~4, 计算所有目标服务的服务效用:
    Service Utility =  $\sum_{i=1}^N \text{term}(i) \times \text{weight}(k)$ 
    term(i) =  $\frac{1}{N} \sum_{i=1}^N \text{Sim}(\text{user}_k, \text{provider}_i)$ 
    其中, weight =  $1/N$ 
    PTN Utility =  $1/2(\text{Consumer Utility} + \text{Provider Utility})$ 
6. 根据服务效用对结果进行评分
7. 选择服务效用最高的结果作为最符合用户需求的云服务
8. 如果用户接受到多个服务方案, 那么用户选择 PTN 效用最高的服务并且拒绝其他服务
结束

```

2.3 本体增强的云服务发现方法的测试结果分析

为了验证本体增强的云服务发现方法的有效性, 设定了系统的测试数据集, 并将测试数据集合存储在 MySQL 数据库中, 用来模拟云服务的注册中心。实验中对 160 个云服务提供者提供的云服务分两种情况进行测试: (1) 不使用云本体的传统方法; (2) 本体增强的云服务发现方法。

在第 1 种情况下, 只有当云服务提供者提供的服务信息与用户的需求信息完全一致时才被认定为查找成功, 许多与用户需求相关的信息被遗漏掉, 降低了服

务的利用率,增加了用户的工作量。在第2种情况下,在本体增强的云服务发现方法中,云本体提供云服务概念、属性及其之间的关系,实现云计算领域知识的重用。利用云本体,使得从基于关键字的服务匹配模式变成了基于语义的服务匹配模式,只要云服务提供者提供的云服务与用户的需求语义相似,相关的云服务就可以被查找出来,从而提升了检索服务的精度,减轻了用户的工作量,很大程度上满足了用户的需求。

实验从以下两个方面对本体增强的云服务方法进行评估:一方面是服务结果的查准率;另一方面是服务结果的查全率。

定义1:设本体增强的云服务发现方法所查找到的服务结果个数为 R ,其中满足用户需求的服务个数为 T ,则服务的查准率如式(4)所示:

$$E = \frac{T}{R}$$

(4)

定义2:设本体增强的云服务发现方法未被查找到的隐含服务结果的总数为 K ,其中与用户需求相关的服务结果个数为 T ,则服务结果的查全率RP如式(5)所示:

$$RP = \frac{T}{K + T}$$

(5)

当对云服务进行精确检索时,未利用云本体的云服务发现方法,网页中的服务信息被随机抽取。如果抽取到的云服务与用户的需求相关,则该云服务的查准率不为零且值唯一。反之,该云服务的查准率为零且查全率也为零。本体增强的云服务发现方法对云服务的描述和用户需求进行相似性匹配,增强了匹配精度,提高了查准率和查全率。

同未利用云本体的云服务发现方法相比,本体增强的云服务发现方法检索到的服务在更大程度上满足了用户的需求。这是因为本体增强的云服务发现方法具有较强的过滤和推理功能。对未利用云本体的云服务发现方法和本体增强的云服务发现方法的查准率和查全率进行分析,得到的相关数据如表3和表4所示。

表3 基于云服务提供者数量的E值比较

云服务提供者	未用云本体的 云服务发现方法	基于云本体的 云服务发现方法
40	0.18	0.76
70	0.08	0.81
100	0.28	0.82
130	0.21	0.79
160	0.43	0.81

图3为两种方法在不同提供者数量下的E值。当云服务提供者数量相同时,未利用云本体的云服务发现方法的查准率低于本体增强的云服务发现方法。

表4 基于云服务提供者数量的RP值的比较

云服务提供者	未用云本体的 云服务发现方法	基于云本体的 云服务发现方法
40	0.21	0.86
70	0.07	0.80
100	0.39	0.89
130	0.33	0.91
160	0.62	0.91

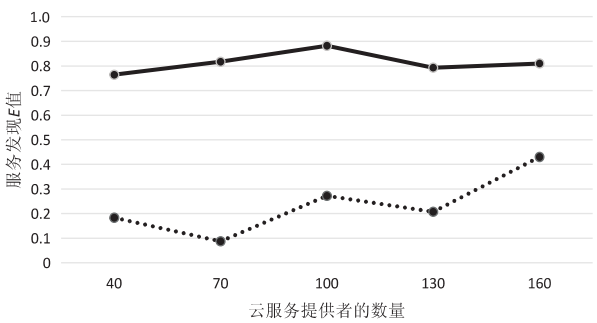


图3 基于云服务提供者数量的E值对比图

图4展示了基于云服务提供者数量的查全率。当云服务提供者数量相同时,本体增强的云服务发现方法的查全率高于未利用云本体的云服务发现方法。

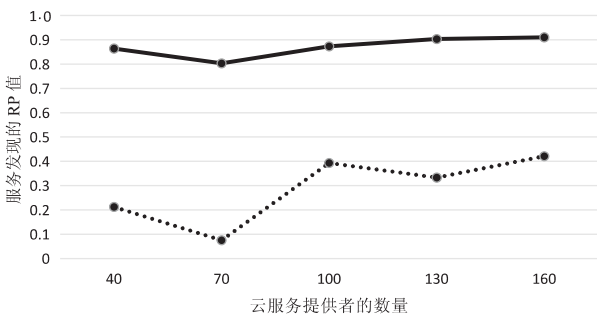


图4 基于云服务提供者数量的RP值对比图

以上分析结果表明,本体增强的云服务发现方法与传统的云服务发现方法相比,提高了服务发现的查准率和查全率,更适合运用于云服务的发现。

3 结束语

针对如何高效地查询满足用户需求的云服务的问题,文中尝试性地提出了本体增强的云服务发现方法。在云服务的发现过程中利用构建的云本体及本体间的概念和属性关系,对云服务的描述信息和用户需求信息进行相似性推理,获得概念之间的相似程度,增强服务的匹配精度,有效地查找到满足用户需求的云服务。

但文中对云环境中服务的发现方法的研究仅处于一个初级阶段,还要做更多的工作。例如,如何对云服务的描述进行统一的规定,怎样从更深的层次构建云

本体,等等。

参考文献:

- [1] Han T, Sim K M. An ontology-enhanced cloud service discovery system[C]//Proceedings of the international multi conference of engineers and computer scientists. [s. l.]: [s. n.], 2010:17-19.
- [2] Kang J, Sim K M. Towards agents and ontology for cloud service discovery[C]//Proceedings of international conference on cyber-enabled distributed computing and knowledge discovery. [s. l.]: IEEE, 2011:483-490.
- [3] Vaquero L M, Rodero-Merino L, Caceres J, et al. A break in the clouds: towards a cloud definition[J]. ACM SIGCOMM Computer Communication Review, 2008, 39(1):50-55.
- [4] Dillon T, Wu C, Chang E. Cloud computing: issues and challenges[C]//Proc of 24th IEEE international conference on advanced information networking and applications. [s. l.]: IEEE, 2010:27-33.
- [5] Kang J, Sim K M. Cloudle: a multi-criteria cloud service search engine[C]//Proc of IEEE Asia-Pacific services computing conference. [s. l.]: IEEE, 2010:339-346.
- [6] 张洁, 何利文, 黄斐一, 等. 一种应用于云计算环境下的服务发现架构[J]. 南京航空航天大学学报, 2013, 45(4):556-562.
- [7] Kang J, Sim K M. Cloudle: an ontology-enhanced cloud service search engine[C]//Proc of web information systems engineeringin. Berlin: Springer, 2011:416-427.
- [8] 宋炜, 张铭. 语义网简明教程[M]. 北京: 高等教育出版社, 2004.
- [9] 黄映辉, 李冠宇. 语义物联网: 物联网内在矛盾之对策[J]. 计算机应用研究, 2010, 27(11):4087-4090.
- [10] Youseff L, Butrico M, da Silva D. Toward a unified ontology of cloud computing[C]//Proc of grid computing environments workshop. [s. l.]: IEEE, 2008:1-10.
- [11] Chapke P P, Keole R R. Ontology of cloud computing[J]. International Journal of Computer Technology & Applications, 2012, 3(3):1189-1193.
- [12] Nagireddi V, Mishra S. An ontology based cloud service generic search engine[C]//Proc of 8th international conference on computer science & education. [s. l.]: IEEE, 2013:335-340.
- [13] Sim K M. Toward an ontology-enhanced information filtering agent[J]. ACM SIGMOD Record, 2004, 33(1):95-100.
- [14] 高岩. 基于本体的信息过滤研究与实现[D]. 沈阳: 沈阳工业大学, 2010.
- [15] Kang J, Sim K M. Cloudle: an agent-based cloud search engine that consults a cloud ontology[C]//Proc of international conference on cloud computing and virtualization. [s. l.]: [s. n.], 2010:312-318.

2016 全国第十四届嵌入式系统学术会议 (ESTC 2016)

2016. 10. 29-30 上海

全国嵌入式系统学术会议 (ESTC) 是由中国计算机学会主办的嵌入式系统专委会年度学术会议, 自 2001 年以来已经成功举办了十三届, 已成为我国嵌入式系统及相关领域的专家、学者、工程师、业界人士以及研究生进行学术交流、技术研讨、产学研互动的重要学术会议。2016 年全国嵌入式系统学术会议将于 2016 年 10 月 29 日~30 日在上海举行。由中国计算机学会嵌入式系统专业委员会、教育部软硬件协同设计技术与应用工程研究中心和华东师范大学计算机科学与软件工程学院共同承办。

ESTC 2016 以“安全可信嵌入式系统设计、验证与应用”为主题。会议旨在讨论嵌入式系统领域的最新研究成果和发展趋势, 开展广泛的学术交流和研讨。ESTC2016 将与 International Symposium on Software and System Reliability (ISSSR) 2016 (<http://www.issr2016.ecnu.edu.cn>) 同期召开, 共享会议特邀报告和论文出版。欢迎从事嵌入式系统及相关领域的专家、学者、工程师、业界人士、研究生踊跃投稿并参加会议。

官网: <http://www.estc2016.ecnu.edu.cn/>

联系人: 曹健 北京大学软件与微电子学院 电话: 010-62769265