

# 基于社会标注的 Web 服务语义自动浮现方法

宁 达 何克清 彭 蓉 冯在文 刘建晓 李 征

(武汉大学软件工程国家重点实验室 武汉 430072)

**摘 要** 社会化标注已经成为当前 Web2.0 时代流行的资源识别和管理方法. 针对当前 Web 服务语义描述能力不足的问题, 提出一种基于多维度的 Web 服务语义社会标注方法. 在社会标注模型的指导下, 利用涉众的广泛参与性, 从推荐标签集、候选标签集以及自由标签集 3 种集合中选取若干个标签对服务进行社会标注; 同时从服务的功能语义、非功能语义、目标语义、交互语义和补充语义 5 个维度建立服务语义社会标注框架, 给出具体标注类型, 将其分为机器标注和群体标注, 利用群体智能, 对机器标注的服务语义进行修正和完善, 提高 Web 服务语义描述的完整性(即服务标签的语义覆盖率). 进而作者提出了一种服务语义自动浮现方法, 结合服务语义维度优先级, 并根据标签出现频率以及服务属性类型优先级对标签进行排序, 使得 Web 服务能够呈现大众用户认可的语义信息, 提高 Web 服务语义描述的准确性(即服务标签的有效使用率), 为后期准确发现 Web 服务奠定基础. 最后通过实验分析, 验证上述方法在提高服务语义描述能力方面的实用性和有效性.

**关键词** 社会标注; Web 服务; 标签过滤; 标签排序; 标签推荐; 语义浮现

**中图法分类号** TP313 **DOI 号**: 10.3724/SP.J.1016.2011.02414

## An Automatic Emerging Approach for Web Service Semantics Based on Social Tagging

NING Da HE Ke-Qing PENG Rong FENG Zai-Wen LIU Jian-Xiao LI Zheng

(State Key Laboratory of Software Engineering, Wuhan University, Wuhan 430072)

**Abstract** Social tagging has become a popular solution for recognition and management of resources in Web2.0 era. To solve the current problem of insufficient capacity of semantic description of Web services, a multi-dimensional social tagging approach for Web services semantics is proposed. Under the guidance of social tagging model, users can choose several tags to annotate services from the three sets of recommendation set, candidate set and free set; simultaneously, social tagging framework of service semantics is established from five dimensions of functional semantics, non functional semantics, objective semantics, interactive semantics and complementary semantics, and specific tagging types given, which are divided into machine's annotation and human's annotation, with the support of collective intelligence, machine's tags can be modified and perfected so that the semantic coverage rate of Web service can be enhanced; Moreover, an approach of service semantics emerges automatically is proposed in this paper, through setting the priority of service semantics dimensions, according to frequency to sort the social tags, so that Web services can show semantic information which the public users approve, and which lays the foundation for later discovering the Web services accurately. Finally, experimental results verify the feasibility and effectiveness of this approach.

**Keywords** social tagging; Web services; tags filtering; tags sorting; tags recommendation; semantic emergence

收稿日期: 2011-08-06; 最终修改稿收到日期: 2011-10-29. 本课题得到国家“九七三”重点基础研究发展规划项目基金(2007CB310801)、国家自然科学基金(60970017)、湖北省杰出青年基金(2009CDA148)、武汉市科技局晨光计划(200950431189)、教育部博士点基金新教师项目(20090141120020)和中央高校基本科研业务费专项基金(武汉大学)(6082008, 3101032, 20102110101000119)资助. 宁 达, 男, 1984 年生, 博士研究生, 主要研究方向为需求工程、服务计算. E-mail: ningdadh@163.com. 何克清, 男, 1947 年生, 博士, 教授, 博士生导师, 主要研究领域为软件工程基础设施中的关键技术. 彭 蓉, 女, 1975 年生, 博士, 教授, 博士生导师, 主要研究领域为需求工程、服务计算. 冯在文, 男, 1980 年生, 博士, 讲师, 主要研究方向为面向服务的计算. 刘建晓, 男, 1984 年生, 博士研究生, 主要研究方向为服务计算. 李 征, 女, 1984 年生, 博士研究生, 主要研究方向为需求工程.

# 1 引言

面向服务的软件开发已经成为目前比较流行的应用系统开发方法,主要是在丰富的网络资源中查找满足用户需求的服务.当前工业界应用广泛的 Web 服务的技术标准(比如 SOAP、WSDL、UDDI 等)主要是从句法角度和静态方式描述服务接口和服务功能,缺乏语义信息,使得 Web 服务的发现和组合只能基于关键字匹配的方式进行,查准率和查全率都不高.因此,语义 Web 服务的概念应运而生.

语义 Web 服务融合了语义 Web 和 Web 服务两大技术,是通过对 Web 服务进行本体语义标识来提高服务的描述能力,使得 Web 服务中每一个信息都具有明确的含义<sup>[1-4]</sup>,由此便于计算机对 Web 服务

进行自动处理和集成.目前存在 3 种将本体用于描述 Web 服务的方法.第 1 种是直接定义一套描述 Web 服务的本体,比如 WSMO<sup>①</sup> 和 OWL-S<sup>②</sup>. WSMO 是利用本体对语义 Web 服务的核心元素定义一个概念模型,OWL-S 是在本体语义 OWL 的基础上进行扩展,用来描述 Web 服务信息.其不足之处是它们仅仅作为 Web 服务描述的上层本体,并不关注特定的应用领域;第 2 种方法是直接在现有的 Web 服务标准(WSDL 和 UDDI)上增加语义信息,采用领域本体直接对 WSDL 文件进行标注,比如 SAWSDL<sup>③</sup>;第 3 种方法是将 WSDL 文档中的元素与本体概念建立对应关系,实现对 Web 服务的语义描述,比如 METEOR-S<sup>[2,4]</sup>.后两种方法虽然能够对 Web 服务进行特定领域的语义标注,但其服务语义描述类型有限.表 1 为 5 种比较典型的语义 Web 服务描述方法.

表 1 相关语义 Web 服务描述的比较

	WSMO	OWL-S	SWSO	SAWSDL	METEOR-S
	Web services	Service	Service	Web service	Service
	Precondition	Input	Input	Input	Input
功能语义描述	Postcondition	Output	Output	Output	Output
	Assumption	Condition	Precondition	—	—
	Effect	Result	Effect	—	—
非功能语义描述	Quality Aspects	ServiceParameter	Service Trust/Reliability	—	—
交互语义描述	Choreography	Orchestration	Orchestration	—	—

上述的 Web 服务语义标注(描述)方法是从 Web 服务提供商的角度考虑如何更好地封装服务以及对服务进行语义描述.然而,Web 服务的提供者和使用者角色一般是由不同企业、不同个人扮演,服务提供者亦根据自身的理解对服务进行封装,无论从设计意图和使用意图是否一致,预期的使用环境和实际使用情境是否相符,开发商提供的服务是否具有更加广泛的用途,还是到设计的 Web 服务是否能被潜在用户发现并使用都存在新的问题,即根据需求如何准确发现已有的 Web 服务成为大家关注的焦点.同时,在查询到具有等价功能的服务中,如何区分它们之间的非功能属性,来满足不同用户的非功能需求,已经成为当前研究的一个瓶颈.为了解决以上问题,大众用户可以根据自己的使用情况对服务资源进行多维度社会标注,挖掘出普遍认同的真实语义,变主观为统计意义上的客观,使得服务呈现更加广泛的语义信息,并通过查询发现与用户期望语义一致的服务资源.

社会标注是大众用户基于个人或者社群的目的,通过自由选择标签来标注网络资源的一种标注行为,不需要遵循一个公共的本体或者词汇表,导致标签质量不高.本文结合语义 Web 服务技术和社会

标注的方法,将社会标注应用于 Web 服务语义描述中,弥补两者各自的不足.针对语义 Web 服务描述文档,由于是服务提供者封装服务,导致服务的描述不能代表用户的意图,所以需要用户的参与,进行社会标注,以展示该服务从用户的角度来看具有的功能;针对社会标注,由于用户具有很少的知识背景以及对服务没有一个系统的认识,需要服务描述文档中的语义信息作为指导,规范其描述服务的方式.因此通过两者的结合,可以弥补彼此之间的不足,使得服务描述更加准确,更加完整.

本文第 2 节介绍 Web 服务语义社会标注方法;第 3 节介绍服务语义自动浮现方法;第 4 节给出实验例子及分析;第 5 节是相关工作的比较;第 6 节是总结和将来的工作.

## 2 Web 服务语义社会标注

### 2.1 服务语义社会标注模型

已有的相关研究工作<sup>[5-9]</sup>在描述服务语义时主

① <http://www.wsmo.org/2004/d10/v0.1/>

② <http://www.daml.org/services/owl-s/1.1/overview/>

③ <http://www.w3.org/TR/2007/REC-sawSDL-20070828/>

要是关注功能语义和非功能语义,对当前云计算环境下以用户为中心的面向服务软件开发具有一定的限制,不能充分体现 Web2.0 时代的特征,亦不能提高用户的体验质量 Quality of Experience (QoE)<sup>①</sup>. 本文对上述的研究工作进行扩展,给出服务语义描述维度框架.

**定义 1.** 服务语义维度,即一级服务语义,可以表示为一个五元组,  $DSS = (F, N, O, I, C)$ . 其中,  $F$  表示服务的功能语义,描述服务的功能属性;  $N$  表示服务的非功能语义,描述服务的质量属性;  $O$  表示服务的目标语义,描述服务所能满足的用户角色和意图;  $I$  表示服务的交互语义,描述服务之间的交互流程;  $C$  表示服务的补充语义,描述服务的补充信息.

**定义 2.** 服务语义社会标注模型是指在用户、标签和服务之间建立的一种多元关系,表示为  $STM = (U, T, S, R \rightarrow)$ , 其中,  $U, T, S$  分别表示  $STM$  中的 3 个基本对象:网络用户、社会标签和服务资源.  $R \rightarrow$  则包含  $STM$  中的 6 个行为关系,如图 1 中的①~⑥.

具体描述如下:

从标注服务的角度看,网络用户试用过该服务资源后通过使用①一组社会标签对服务资源进行标注②,来展示对该资源的理解;或者选择①已经存在的社会标签对该服务资源进行标注,展示对该标签来标注服务的支持和认同;同时,在标签达到一定规模时,通过自动推荐④,使得用户能够根据推荐的结果从海量标签库中选择符合自身期望的标签来标注服务资源. 从使用服务的角度看,用户能够对试用(使用)过的服务进行投票或者评价③,以展示用户对该服务资源的认可度,同时服务资源也能够通过推荐⑤)机制返回给用户,供用户选择以满足自身的个性化需求. 此外也能够通过对使用过相同服务或标签的用户进行推荐⑥,以展示相似用户的行为轨迹和偏好. 本文主要关注①~④方面的内容(即图中的实线关系).

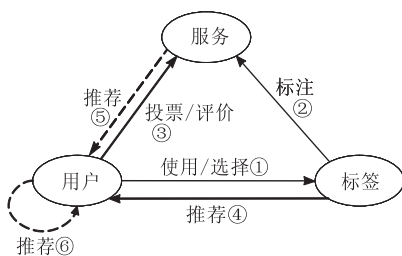


图 1 服务语义社会标注模型

## 2.2 社会标注方式

**定义 3.** 给定一个服务  $S$ , 其候选标签集为各个服务语义维度的集合形式,表示为  $CTS = \{F, N, O, I, C\}$ . 其中,  $F \cap N \cap O \cap I \cap C = \emptyset$ ,  $|F| + |N| + |O| + |I| + |C| = n$ ,  $|F|, |N|, |O|, |I|, |C|$  分别为各语义维度集合中标签的个数.

**定义 4.** 给定一个服务  $S$ , 其推荐标签集  $RTS = \{F_i, N_j, O_k, I_l, C_m\}$ . 其中,  $F_i, N_j, O_k, I_l, C_m$  分别为各语义维度集合中的某个标签.  $1 \leq i \leq n, 0 \leq j, k, l, m < n$ .

**定义 5.** 自由标签集是指根据用户的知识背景和认知水平所构成的一种虚拟的标签集合,用  $FTS$  表示.

对一个服务资源而言,假设已经存在一定数量的标签,可以通过浏览该资源的标签并进行标注活动. 用户可以选择其中已经存在的标签(图 2 中的实线圆角矩阵)作为对该资源的一种认知或理解;如果标签过多,用户无法浏览所有标签时,可以根据自身对该资源的理解对其进行标注(图 2 中的虚线圆角矩阵). 因此用户对服务进行标注的方式,主要包含以下 3 种:

(1) 通过推荐集进行标注. 这种方式是通过读取用户的个人信息自动从标签库中推荐 5 个不同服务语义维度中最能满足用户期望的标签,是一种个性化标签推荐方法.

(2) 通过候选集进行标注. 这种方式是通过从标签库中某个服务的所有标签按照服务语义维度优先级(见定义 9)、服务语义属性类型优先级(见定义 10)以及标签出现频率进行加权排序,将排序后的标签集展现给用户,以供选择,是一种大众化标签推荐方法.

(3) 通过自由标签集进行标注. 这是一种最基本的标注方式,也是服务资源的初始标注方式. 用户不必受已存在标签的影响,而是根据自身对服务资源的理解,利用标注模板对服务资源进行标注,这也是提高标签语义覆盖率(见定义 7)最直接最有效的途径. 该部分我们将在下节重点讨论.

用户对服务资源进行标注的方式以及各方式之间的关系如图 2. 该图中,用户使用 3 种方式来标注服务资源,其中自选标签集为虚集,依赖于用户所具有的知识背景以及对事物的认知,通过使用该集合

① [http://en.wikipedia.org/wiki/Quality\\_of\\_experience](http://en.wikipedia.org/wiki/Quality_of_experience)

中的标签可以将其实化为候选标签. 候选标签集和推荐标签集为实集, 即存在的真实的标签集合. 候选标签集为服务资源的所有标签集合, 推荐标签集为候选标签集的子集, 主要是通过提取用户的个人信息, 从候选标签集中甄选出符合用户个性化需求的标签.

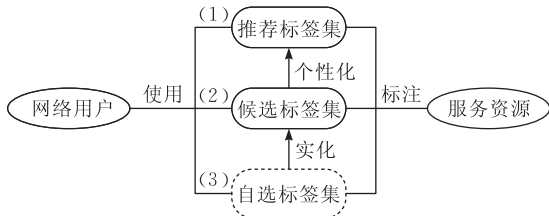


图 2 社会标注方式

### 2.3 服务语义社会标注类型

本文着重从机器标注(MA)和群体标注(HA)两个方面进行讨论. 机器标注是从服务提供商和服务运营商的角度自动监测获取到的服务语义标签, 同时也可以由用户进行修正和完善. 主要包括从服

务描述文档(比如 WSDL、OWL-S、BPEL4S 等)中提取的操作(Operation)、消息(Message)、输入输出参数(⟨Input, Output⟩)、交互流程(Interaction)、服务所归属的领域(Domain)以及从服务运营商获取的服务质量<sup>[9]</sup>, 包括响应时间(Response Time)、可靠性(Reliability)、可用性(Availability)、吞吐量(Throughput)、安全性(Security)、服务价格(Price)、提供商的名称(Name)、所归属的国家(Country)以及服务推荐级别(User Rating)等方面的标注; 群体标注是从服务消费者的角度考虑根据用户的知识背景以及推荐的标签对服务语义进行手动标注. 主要包括在具体情境下使用服务后对服务的实体(Entity)描述、特征(Characteristic)描述、对服务的评价(Evaluation)以及用户所扮演的角色(Role)和用户的意图(Intent)等方面的标注. 同时, 用户也能够对机器标注的服务属性进行修正、完善和评价, 以展示哪些标签更加符合自己的意图和期望. 如图 3.

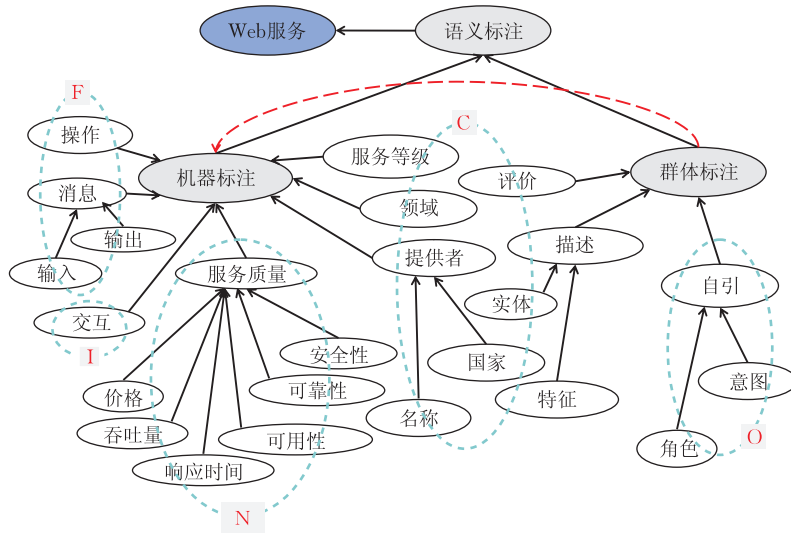


图 3 服务语义社会标注类型

**定义 6.** 服务语义属性类型, 即二级服务语义, 表示为  $TSSP := \langle Ope, Inp, Out, Sec, Rel, Ava, Res, Thr, Pri, Rol, Inte, Int, Dom, Nam, Cou, Ent, Cha, UR, Eva \rangle$ . 由输入信息、输出信息、安全性、可靠性、可用性、响应时间、吞吐量、价格、角色、意图、领域、服务提供商名称、国家、实体、特征、操作、交互流程、服务推荐级别以及评价构成的服务描述信息.

**定义 7.** 服务标签的语义覆盖率指能够准确、完整描述服务语义信息的标签类型数所占整个服务语义标签类型数的比例, 表示为  $SCR = (|MT| + |HT|) / |TSSP|$ . 其中,  $|MT|$  指机器标签类型数,

$|HT|$  指人工标签类型数. 本文仅考虑服务语义一级覆盖率, 即本文所指的是服务语义覆盖率. 对各个标签类型内部的语义覆盖率(即服务语义二级覆盖率), 目前本文暂不考虑. 文中将服务语义属性类型作为社会标注语义覆盖率的参考标准, 当用户完成服务语义的所有标注类型时, 我们认为该服务的语义覆盖率达到 100%.

**定义 8.** 服务标签的有效使用率指能够体现服务语义的标签数目占该服务所有标签总数的百分比, 表示为  $ER = |Semantic\_Tags| / |Total\_Tags|$ . 其中,  $|Semantic\_Tags|$  指能够体现服务语义的标

签数,  $|Total\_Tags|$  为服务的标签总数.

用户查询服务时,能够根据偏好从不同层面对已经查询到的服务进行过滤和排序,呈现最能满足其个性化需求的服务.服务语义属性优先级的设定可以为服务的自动呈现提供一种快速准确的解决方案.包括服务语义维度优先级  $PDSS$  和服务语义属性类型优先级  $PTSSP$ .

**定义 9.** 服务语义维度优先级  $PDSS$ . 针对一级服务语义,由于服务的功能属性  $F$  是用户必须实现的,因此具有最高优先级;非功能属性  $N$  是对功能属性的约束,具有次优先级;目标属性  $O$  是服务能够满足用户所扮演的角色需求,具有第三优先级;交互属性  $I$  是描述服务之间的交互流程,能够对组合服务进行控制,具有第四优先级;而补充属性  $C$  是对服务属性的补充说明,具有最低优先级.因此这五者之间的优先级排序可表示为  $PDSS \llbracket F, N, O, I, C \rrbracket$ , 即  $F \triangleright N \triangleright O \triangleright I \triangleright C$ ,  $\triangleright$  表示弱序关系.

**定义 10.** 服务语义属性类型优先级  $PTSSP$ . 针对二级服务语义,由于各个一级服务语义维度内部所包含的各个属性类型具有同等优先级,因此需要根据用户历史信息(或者根据用户关注点的不同)进行优先级动态设定.

**定义 11.** 用户历史信息或用户关注点. 用户

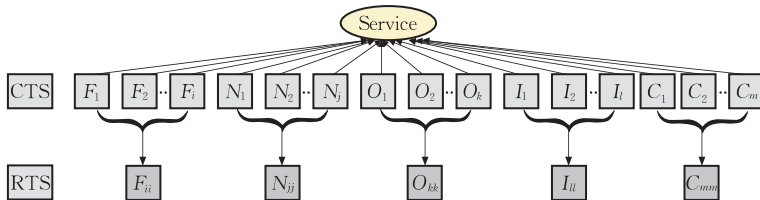


图 4 服务的 CTS 和 RTS

**定义 13.** 标签排序值和出现频率. 在  $t$  时刻,对一个服务  $S$ , 有  $n$  个标签对其描述, 表示为  $S = \sum_{i=1}^n \{CS_i\}$ ,

标签  $CS_i$  的使用次数为  $C_i$ , 排序值为  $V_i$ ,  $\sum_{i=1}^n V_i = 1$

$$V_i = C_i / \sum_{i=1}^n C_i \quad (1)$$

标签  $CS_i$  的出现频率

$$f_i = C_i / (t_{last} - t_{first}) \quad (2)$$

### 3.2 标签过滤

服务语义社会标签过滤的目的是为了提高社会标签的有效使用率,即服务语义描述的准确性.

**算法 1.** 服务语义社会标签过滤算法 STFSS.

输入: 用户输入的标签集  $\{t_i\}$ ; 词汇库  $WORDS$ , 修饰词库  $DECORATION$ , 无意义词库  $NONSENSE$ ,

历史信息包括静态信息 ( $staticInfo$ ) 和动态信息 ( $dynamicInfo$ ). 静态信息主要是描绘用户自身的一些静态特征, 比如社会角色 ( $Social\_Role$ )、职业 ( $Career$ )、偏好 ( $Preference$ ) 等; 动态信息主要是描述用户访问过的服务以及使用过的标签.

## 3 服务语义自动浮现

### 3.1 问题描述

针对当前网络中丰富的服务资源  $Service$ , 使用社会标注方法对其进行语义描述, 如何保证服务标签的准确性以及标签推荐的合理性, 以下算法将为其提供解决方案.

**定义 12.** 服务  $S$  的  $CTS$  和  $RTS$  实例. 如图 4, 给定一个已发布的服务  $Service(S)$ , 存在  $n (n \geq 1, n \in Z)$  个标签对其进行标注. 其中,  $\{F_1, F_2, \dots, F_i\}$ ,  $\{N_1, N_2, \dots, N_j\}$ ,  $\{O_1, O_2, \dots, O_k\}$ ,  $\{I_1, I_2, \dots, I_l\}$ ,  $\{C_1, C_2, \dots, C_m\}$  分别为功能语义  $F$ 、非功能语义  $N$ 、目标语义  $O$ 、交互语义  $I$  以及补充语义  $C$  5 种属性维度的标签集实例,  $i+j+k+l+m=n$ ;  $F_{ii}$ ,  $N_{jj}$ ,  $O_{kk}$ ,  $I_{ll}$ ,  $C_{mm}$  分别为每个维度的推荐标签实例,  $ii \in (1, i)$ ,  $jj \in (1, j)$ ,  $kk \in (1, k)$ ,  $ll \in (1, l)$ ,  $mm \in (1, m)$ .

语义词汇库  $WORDNET$ ; 自动提取的服务  $S$  语义标签集  $\{MA_i\}$

输出: 过滤后的标签集  $\{t'_i\}$

1. begin
2.  $\{t'_i\} \leftarrow \{MA_i\}$ ; // 初始化标签集
3.  $MA_i.count \leftarrow 1$ ; // 初始化机器标签使用次数
4.  $FilterTagsSet \leftarrow nil$ ; // 初始化过滤标签集
5. input  $\{t_i\}$ ;
6. for  $i=1$  to  $i=|\{t_i\}|$  do
7. if  $t_i \in WORDS$  // 判断输入标签是否为合法单词
8. if  $t_i \in DECORATION \parallel t_i \in NONSENSE$   
// 判断输入标签是否为修饰词或无意义词
9.  $FilterTagsSet \leftarrow t_i$ ;  
// 将输入标签添加到过滤标签集中
10. delete  $t_i$ ;
11. else
12. if  $t_i \in \{t'_i\}$  // 判断输入标签是否已经存在

```

13.    $t_i.count++$ ;
14.   elseif WORDNET( $t_i, t'_i$ ) //判断输入标签是否
      //和存在标签为同义或近义,并记录关联
15.   if synonym( $t_i, t'_i$ )
16.     record  $t_i.R.synonym t'_i$ ;
17.   else similar( $t_i, t'_i$ )
18.     record  $t_i.R.similar t'_i$ ;
19.   add  $t_i$  to  $\{t'_i\}$ ; //将输入标签添加到服务标签集
20.   end if
21. end if
22. end if
23. else
24.   FilterTagsSet  $\leftarrow t_i$ ;
25.   delete  $t_i$ ;
26. end if
27. end for
28. |Semantic_Tags|  $\leftarrow \{t'_i\}.count$ ; //计算服务标签数目
29. |Total_Tags|  $\leftarrow FilterTagsSet.count + \{t'_i\}.count$ ;
      //计算过滤标签数目
30. ER = |Semantic_Tags| / |Total_Tags|;
      //计算服务标签有效率
31. return  $\{t'_i\}$ ; //输出过滤后的标签集
32. end

```

算法 1 的基本思路是,当用户输入标签集时,首先检查该标签是否为不合法、修饰以及无意义单词,如果是,则将该标签添加到过滤标签列表中;否则通过语义库判断该标签是否为已存在标签的同义或相似标签<sup>[10]</sup>,如果是,则记录它们之间的关联,将该标签添加到服务标签库中.然后根据过滤标签数目和服务标签库中的标签数目来计算服务语义社会标签的有效率.最后输出过滤后的标签列表.

### 3.3 标签排序

标签排序的目的是为了让用户能够根据标签的优先关系了解该服务的语义以及提高服务语义描述的完整性.

**算法 2.** 服务语义社会标签排序算法 STSSS.

输入: 过滤后的标签集  $\{t'_i\}$ ; 机器自动提取的服务语义标签集  $\{MA_i\}$ ; 服务语义维度 DSS; 初始化推荐标签集  $\{s_i\}$

输出: 排序后的服务语义标签集  $\{t'_i\}$

```

1. begin
2. |TSSP|  $\leftarrow 19$ ;
3. |MA|  $\leftarrow \{MA_i\}.count$ , |HA|  $\leftarrow nil$ ;
4. input  $\{t'_i\}$ ;
5.  $\{t'_i\} \leftarrow \{t'_i\}$ ;
6. for  $i=1$  to  $i=|\{t'_i\}|$  do
7.   if  $t'_i \in Func\_S$ 
8.     add  $t'_i$  to Func_S; //将  $t'_i$  添加到集合 Func_S 中

```

```

9.   Func_S.count; //计算 Func_S 中的标签个数
10.  if  $t'_i.type \in MA_i$ 
11.    |MA.type|.count++;
12.  else
13.    |HA.type|.count++;
14.  end if
15.  ..... //判断  $t'_i$  是否属于 NFunc_S, Obj_S, Int_S,
      //或者 Comp_S, 步骤如 7~14
16. end if
17.  $C_i \leftarrow t'_i.count$ ;
18.  $f_i = C_i / (t_{last} - t_{first})$ ; //计算  $t'_i$  的出现频率
19.  $V_i = C_i / (C_1 + C_2 + \dots + C_n)$ ;
20. end for
21. input [F, NF, O, I, C]; //给定服务语义维度优先级
22. rank( $f_i, f_j$ ); //对标签的出现频率进行排序
23. CS  $\leftarrow \{t'_i\}.rank$ ;
      //结合服务语义维度优先级对  $\{t'_i\}$  进行排序
24. SCR = (|MT| + |HT|) / |TSSP|;
      //计算服务语义的覆盖率
25.  $\{t'_i\} \leftarrow CS$ ;
26. return  $\{t'_i\}$ 
27. end

```

算法 2 的基本思路是,将过滤算法的输出作为标签排序算法的输入,根据服务语义维度 DSS 对标签进行分类,按照服务语义维度优先级 PDSS 进行排序,并根据标签的出现次数在各个维度类型中计算标签的出现频率和标签排序值,如果标签出现次数大于 1,则根据式(2)计算标签的出现频率,否则,根据式(1)计算标签排序值.标签出现次数大于 1 的标签,其排列顺序优先于出现次数为 1 的标签.最后根据机器标签类型和人工标签类型计算服务语义覆盖率.

### 3.4 标签推荐

标签推荐是根据用户的不同属性信息或者用户不同的关注点从海量服务标签库中甄选出 5 个不同维度的标签推荐给用户,目的是能够让用户快速选择适合自身需求的标签进行标注活动.

**算法 3.** 服务语义社会标签推荐算法 STRSS.

输入: 排序后标签集  $\{t'_i\}$ ; 用户信息 userInfo; 用户信息与服务语义属性类型对应关系表 RELATIONSHIP

输出: 推荐标签集  $\{s_i\}$

```

1. begin
2.  $\{s_i\} \leftarrow nil$ ;
3.  $\{s_i\}.COUNT \leq 5$ ;
4. for  $i=1$  to  $i=|\{t'_i\}|$  do
5.   if  $t'_i \in Func\_S$ 

```

```

6.   $userInfo \leftarrow t'_i.user.staticInfo$  and
     $t'_i.user.dynamicInfo$ ;
7.  record the  $userInfo$ ;
    //记录用户的静态信息和动态信息
8.   $RELATIONSHIP(userInfo, TSSP)$ ;
    //使用用户信息与服务语义属性类型关联表
9.   $t'_i.user.userInfo.RELATIONSHIP.TSSP$ ;
    //根据关联表查找出该用户对应的
    //服务语义属性类型
10.  $t'_i.rank \leftarrow TSSP.rank$ ; //对服务语义属性类型进行
    //排序,即对相应的标签进行排序
11.  $F_{ii} \leftarrow t'_i.rank$ ; //根据用户信息推荐标签
12.  $\dots$  //依次进行  $NFunc\_S, Obj\_S, Int\_S,$ 
    //或  $Comp\_S$  的判断
13. end if
14. end for
15.  $RS \leftarrow \{F_{ii}, N_{jj}, O_{kk}, I_{ll}, C_{mm}\}$ ;
16.  $\{s_i\} \leftarrow RS$ ;
17. return  $\{s_i\}$ 
18. end

```

算法 3 的基本思路是,判断排序后的标签属于哪种服务语义维度,在该服务语义维度内,提取该标签的用户信息,并查找用户信息与服务语义属性类型对应关系表,提取该用户最相关的标签作为推荐标签.如果同时存在两种类型,则计算各自的标签出现频率,取最大者即为推荐标签.

### 3.5 服务语义自动浮现

服务语义自动浮现是一个根据用户信息、用户标注等不断反馈、不断更新的活动.用户在对服务进行标注时,根据用户信息可以自动提取满足其需求的推荐标签;如果该推荐标签不能满足用户需求,用户能够从候选标签集中选择满足其需求的标签进行标注,也能够自行添加标签,整个标注活动能够使服

务语义描述处于一种不断完善的状态.因此,上述标签过滤、标签排序以及标签推荐 3 种算法并不是孤立存在的,它们之间相互依赖,相互协作,进而实现服务语义的自动浮现.

在整个服务语义自动浮现算法中,将分两种情况进行讨论.第 1 种情况是从  $CTS$  和  $RTS$  中选择标签,该过程不需要进行标签过滤,而是在标签排序和标签推荐中将用户的标注行为看成是对某个标签的认可,使得该标签在服务语义描述中的权重增加,能够呈现大众用户认可的语义信息,该过程能够提高服务语义标签的有效使用率;第 2 种情况是从  $FTS$  中选择标签,该过程需要进行标签过滤,完成后的过程与第 1 种情况类似.然而该过程却是提高服务语义覆盖率的基本途径.

## 4 实验分析

### 4.1 数据集

为了对实验的可迁移性进行验证,本文首先从公共平台“服务搜索引擎 seekda”中爬取了 5910 个 Web 服务,选择使用较多的 10 个服务作为标注案例,并抽取相应的服务标签集作为数据集 1,采集时间为 2011 年 6 月 4 日,见表 2;同时将这些服务移植到作者所在的武汉大学群体智能研究组开发的测试平台 SKLSEWiki 之中.实现环境包括:(1)硬件:2.53 GHz CPU,4 GB 内存,320 GB 硬盘,网络适配器,Internet 环境;(2)软件:Windows7 操作系统,IE7 浏览器,MySQL 5.1.33 数据库服务器,Apache2.2.11 应用服务器,PHP5.2.9-2 服务端运行时环境;(3)编程语言:PHP、JavaScript、HTML、CSS.

表 2 seekda 中采集的 10 个服务及服务对应的标签

服务	标签
S1:GlobalWeather	weather, global, company, city, current, <i>MyWeather</i> , condition, aviation, <i>mywebservice</i> , <i>worldwide</i> , free
S2:CurrencyServerWebService	free, currency, currency conversion, exchange rates, company
S3:ValidateEmail	address, email, email server, <i>email</i> , <i>emailVerification</i> , company, validation
S4:country	<i>currency</i> , company, code, <i>Greenwich mean time</i> , name, <i>international dialing code</i> , currency name, GMT, currency code, country, free
S5:StockQuote	stock, company, stock quote, symbol, finance
S6:TranslateService	Korean, Italian, French, Spanish, Japanese, English, translation, language, Portuguese, <i>Russion</i> , German, Chinese, company
S7:WSNewHotelSrv	hotel, internal, sign-up, new, <i>hotel</i> , <i>sign-up</i> , <i>internal</i>
S8:airport	latitude, country, <i>degree</i> , minute, airport, code, <i>feet</i> , name, runway length, second, company, abbreviation, GMT, elevation, longitude
S9:LocationByZipService	<i>geocode</i> , coordinate, distance, latitude, free, Zip code, calculation, USA, longitude
S10:MathService	substraction, addition, example, number, <i>community</i> , multiplication, free, mathematics, division, <i>mathematics</i> , <i>unknown</i> , <i>sample</i>

**实验 1.** 服务语义社会标注的可迁移性。  
随机选择 5 个用户在 SKLSEWiki 平台上对上述 10 个服务进行社会标注, 将服务和

服务对应的标签作为数据集 2, 采集时间为 2011 年 6 月 10 日, 见表 3(说明: 这 5 个用户事前并未学习服务语义社会标注类型)。

表 3 SKLSEWiki 中采集的 10 个服务及服务对应的标签(未使用标注类型)

服务	标签
S1: GlobalWeather	weather, business, global, free, country name, cities, <i>weather</i> , climate, <i>travel</i> , <i>air temperature</i> , wind scale, slowly-speed
S2: CurrencyServerWebService	dollar, current, rate, <i>unknown</i>
S3: ValidateEmail	email address, validate, company, <i>validation</i> , <i>boolean</i> , <i>email</i> , <i>email address verification</i> , <i>different protocol service</i>
S4: country	china, Japanese, Chongqing, Washington, free, name, <i>code</i> , country, country code, currency code, <i>country name</i> , <i>international dialing code</i>
S5: StockQuote	stock, NASDAQ, <i>Hang Seng</i> , symbol, finance, <i>stock quote</i> , company, stock quotation, stock market, <i>complicated operation</i>
S6: TranslateService	translation, Chinese, English, language, paper, <i>language</i> , <i>translating</i> , convert, online translation, <i>language translation</i> , multi-language
S7: WSNewHotelSrv	hotel, trip, internal, sign-up, <i>hotel</i> , prices, ReservationsInformation, RoomsTypes, <i>unknown</i> , <i>teste</i>
S8: airport	chongqing, wuhan, Hawaii, runway_length, minute, <i>degree</i> , airport code, country, CountryCode, airport information, <i>airport code</i>
S9: LocationByZipService	zip, <i>wuhan university</i> , longitude, distance, <i>zip code</i> , <i>geocode</i> , geographic location, worldwide
S10: MathService	statics, probability, matrix, mathematics, number, <i>sample</i> , addition, <i>add</i> , divide, multiply, subtract, online calculator, <i>compute</i>

针对以上 10 个服务, 本文从标签总数、标签类型数、重复标签数、无意义标签数以及错误标签数 5 个维度来验证从数据集 1 到数据集 2 的可迁移性, 维度值  $D_1$  和  $D_2$  取 10 个服务中相应维度最小值以及最大值的并集。以维度“标签总数”为例, 在数据集 1 中, S5 中的标签个数最少(5 个), S8 中的标签个数最多(15 个), 因此取  $D_1 \in [5, 15]$ 。统计表 2 和表 3 中的数据(其中斜体部分表示不能反映服务语义的标签), 可以得到表 4。通过对表 4 的分析, 可以看出数据集 1 和数据集 2 在如下几个方面存在一些不足: (1) 冗余以及无实际意义的社会标签众多(如 weather, unknown 等); (2) 社会标签仅仅反映了服务的一些特征以及该服务是什么, 并没有反映该服务能做什么, 即没有体现该服务实现了哪些功能和非功能需求; (3) 社会标签未分类, 排序较混乱, 导致标签类型数较低。因此不能体现服务的真实语义, 也不能为其它用户了解或使用该服务提供有效地指

导。同时由于数据集 1 和数据集 2 中的各个维度值相近, 可以分析得出 SKLSEWiki 平台中未使用服务语义社会标注类型的用户和 seekda 网站中的用户对服务标注的标签没有太大差别。因此说明该实验具有可迁移性。

表 4 数据集 1 和数据集 2 的可迁移性比较

	数据集 1 ( $D_1$ )	数据集 2 ( $D_2$ )
标签总数	$5 \leq D_1 \leq 15$	$4 \leq D_2 \leq 13$
标签类型数	$2 \leq D_1 \leq 4$	$2 \leq D_2 \leq 5$
重复标签数	$0 \leq D_1 \leq 3$	$0 \leq D_2 \leq 3$
无意义标签数	$0 \leq D_1 \leq 2$	$0 \leq D_2 \leq 3$
错误标签数	$0 \leq D_1 \leq 1$	$0 \leq D_2 \leq 1$

## 实验 2. 服务标签的语义覆盖率和有效使用率。

随机选择不同于实验 1 中的 5 个用户使用本文提出的社会标注方法在 SKLSEWiki 平台上对上述 10 个服务进行标注, 将服务和服务对应的标签作为数据集 3, 采集时间为 2011 年 6 月 17 日, 见表 5。

表 5 SKLSEWiki 中采集的 10 个服务及服务对应的标签(使用标注类型)

服务	过滤后的标签	被过滤的标签
S1: GlobalWeather	GetCitiesByCountry, GetWeather, CountryName, CityName, Availability-87.27%, traveler, get weather information, weather, fast, current, good, city, date, temperature, webservicex.com	<i>placename</i> , <i>good</i>
S2: CurrencyServerWebService	getCurrencyValue, getDataSet, getDollarValue, getProviderDescription, getProviderList, getProviderTimestamp, getXnlStream, provider, currency, Availability-98.72%, worker, slowly, finance, currencyserver.de, Germany	<i>student</i> , <i>currency</i>
S3: ValidateEmail	IsValidEmail, Email, boolean, student, fast, Availability-87.92%, webservicex.com	<i>email address</i>

(续 表)

服务	过滤后的标签	被过滤的标签
S4 :country	GetCountries, GetCountryByCountryCode, GetCountryByCurrencyCode, GetCurrencies, GetCurrencyByCountry, GetCurrencyCode, GetCurrencyCodeByCurrencyName, GetGMTbyCountry, GetISD, GetISOCountryCodeByCountryName, CountryName, CountryCode, International Dialing Code, Availability-85.41%, teacher, reporter, travel, city, country, slow, free, webservicex. net	country
S5 :StockQuote	GetQuote, symbol, Availability-87.49%, buyer, quote, slow, bad, webservicex. com, finance	slow, unknown
S6 :TranslateService	Translate, Text, LanguageMode, languages, Availability-87.23%, clerk, do some translation, slow, school, bad, webservicex. com	translation
S7 :WSNewHotelSrv	AddAllotmentPlan, AddGroup, AddHotel, AddUser, ChangeAdministratorPassword, DeleteGroup, DeleteHotel, DeleteUser, GetAllGroups, GetAllHotels, GetAllUser, Availability-99.15%, traveler, get hotel information, country, place, hotel, price, traffic, slow, bad, newhotel. com, Portugal	myservice, slow-speed
S8 :airport	GetAirportInformationByCountry, GetAirportInformationByAirportCode, GetAirportInformationByAirportName, GetAirportInformationByISOCountryCode, country, airportCode, CountryAbbreviation, Availability-89.04%, person, get airport information, city, airplane, airport, practical, transport, good, free, webservicex. com	
S9 :LocationByZipService	getDistanceByZip, getLocationByZip, zip, Availability-68.47%, traveler, get location information, location, country, slow, transport, bad, flash-db. com, USA	
S10 :MathService	Add, Divide, Multiply, Subtract, Availability-99.40%, student, worker, clerk, minus, math, fast, education, good, aspalliance. com, USA	adding

表 5 是用户使用本文的社会标注方法所生成的数据集,可以看出使用标签类型可以减少无意义和错误标签数,并根据本文的服务语义自动浮现方法,可以对标签进行有效排序,使得服务的语义能够展现.图 5 为未经过标签分类(即未使用服务语义标注类型)的服务(TNCS)的标签个数,其中包括有效标签(ET)数和过滤标签(FT)数.图 6 为经过标签分类(即使用服务语义标注类型)的服务(TNCS)的标

签个数.经过对比可以看出,TCS 中的服务标签总数和 ET 个数多于 TNCS 中的 ET 个数,同时 TCS 中的 FT 个数明显少于 TNCS 中的 FT 个数.针对图 7 中的服务标签类型数,TCS 中的标签类型数则明显多于 TNCS 中的标签类型数.

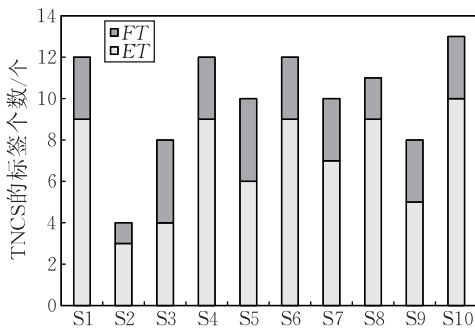


图 5 未经过标签分类的服务标签数目

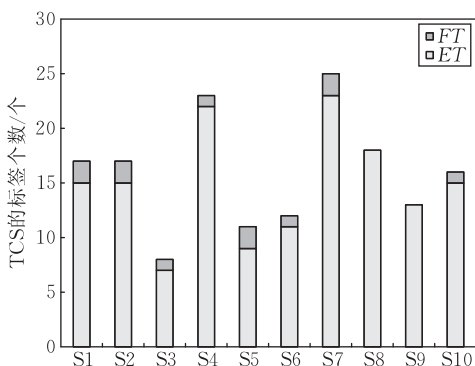


图 6 经过标签分类的服务标签数目

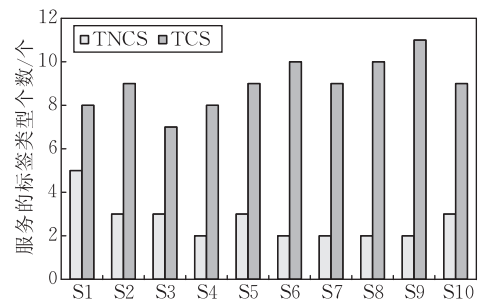


图 7 服务的标签类型数目

## 4.2 结果分析

通过对数据集 2 和数据集 3 的对比分析,没有使用标签分类的服务,其语义信息相对贫.然而,使用基于标签分类的社会标注方法,服务标签的语义覆盖率相对丰富.在用户对服务进行标注前,服务已经存在自动提取的机器标签,使得服务在开始阶段就具有一定的语义信息.而且当标注行为一经发生时,语义信息覆盖率将快速提高,这里我们使用 RSS/Atom feed 技术同步监测该服务资源,该服务资源的信息一旦有所更新,则立即通知相关用户,用户可以浏览该服务资源,或者使用或者对该服务资源进行描述,如图 8.而对服务标签的有效使用率而言,在没有使用标签分类时,由于没有规范的指导,用户在标注服务时显得比较随意,导致服务标签的

有效使用率较低;然而,在使用标签分类体系后,用户能够比较清楚地了解服务语义的标注类型,因此在标注服务时就具有一定的方向性和目的性,使得服务标签具有较高的有效使用率,如图 9.

由此分析得出,用户使用本文提出的方法将大大提高服务语义的覆盖率,同时服务描述的准确性也得到相应提高,将有助于后期的服务发现.

### 实验 3. 服务语义社会标注的可实践性.

本文从用户标注和服务发现两个角度说明服务语义社会标注方法具有可实践性.

(1)从服务发现的角度考虑,为了说明该方法在提高服务发现能力方面具有一定的优势,本实验从服务的查准率和平均查找时间方面分别在 seekda 和 SKLSEWiki 平台上对上述 10 个服务进行发现,见表 6. 由于本文重点关注的是服务语义社会标注和语义自动浮现方法,因此仅仅通过多维度的关键词匹配的方式进行查找,更加精确的语义查询将另文介绍.

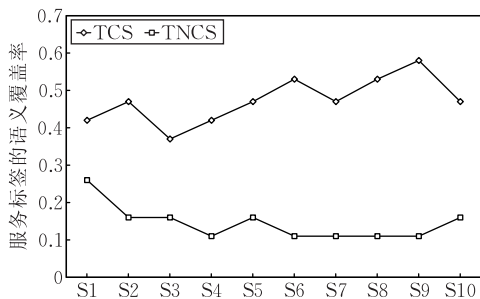


图 8 服务标签的语义覆盖率

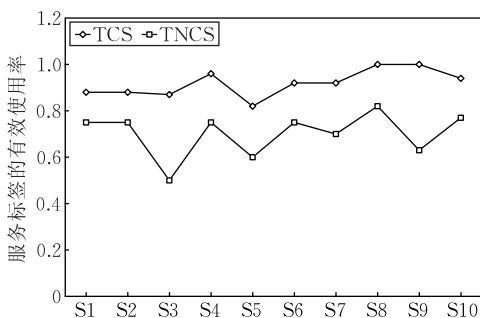


图 9 服务标签的有效使用率

表 6 服务语义社会标注可实践性比较 1

	查准率	平均查找时间/s
seekda 平台	4/10	1.2
SKLSEWiki 平台	8/10	0.9

(2)从用户标注的角度考虑,为了说明该方法能够被用户所认可,本实验对上述 10 个服务的标注时间以及 10 个用户是否愿意使用该方法标注服务进行了统计,见表 7.

表 7 服务语义社会标注可实践性比较 2

	标注时间	是否意愿	
实验 1	U1	8	是
	U2	5	是
	U3	7	是
	U4	10	否
	U5	7	是
实验 2	U6	30	否
	U7	25	是
	U8	27	是
	U9	20	否
	U10	60	是

从表 7 可以看出,用户在使用标签类型来标注服务的同时,虽然服务标签的语义完整性和准确性得到提高,但却消耗了较多的标注时间. 对网络用户的意愿程度进行分析可以看出大多数用户还是比较愿意使用本文所提出的标注方法.

## 5 相关工作比较

针对当前涉及到的服务语义社会标注方法,主要从以下两个角度展开对比研究.

第 1 方面: 网络资源的社会标注方法.

针对当前涉及到的网络资源的社会标注方法,业界已有相关研究. 近年来,出现了一些网络资源的社会标注网站,如 seekda<sup>①</sup>, dil. icio. us<sup>②</sup> 等等. 国际和国内的相关学者在社会标注系统(dil. icio. us)的基础上展开研究. IBM 中国研究院张俐<sup>[11]</sup>等研究员对 Web2.0 环境下的社会标注进行了一定的阐述. Vojnovic 等人<sup>[12]</sup>提出一种基于用户反馈的资源项排序与推荐方法,根据用户偏好,建立用户选择模型,实现对资源项的优化排序;中国科学院计算技术研究所史忠植研究员提出一种层次聚类方法<sup>[13]</sup>,利用网络资源和它们的标签之间的关系对网络资源进行聚类;其他研究者把用户及其使用的标签作为一种社会网络<sup>[14-15]</sup>,研究用户的行为趋势,从而提高资源查询效率.

然而,上述方法仅仅将标签、资源和用户之间的关系作为研究对象,并不关注标注行为,导致标签的质量不高. 本文的方法不仅研究了以上 3 个对象的关系,也对标注行为进行了深入研究,能够为用户提供一个表述规范、语义完整的标注信息,从而为后期精确发现资源提供指导.

第 2 方面: Web 服务语义标注(描述)方法.

① <http://webservices.seekda.com/>

② <http://www.delicious.com/>

为了对 Web 服务语义信息进行描述,清华大学李涓子<sup>[16]</sup>提出一种 Web 服务语义标注方法,以现有的 Web 服务描述语言(WSDL)文件为基础,充分利用 WSDL 中 XML schema 的语义信息与本体实体的相似性,从 WSDL 文件中抽取语义信息,并利用实体之间的名称相似度与结构相似度进行语义标注,生成基于 OWL-S 的 Web 服务语义描述. 同济大学的张大陆等人<sup>[17]</sup>提出一种基于概念频率的 Web 服务语义标注方法,通过得到概念的不同属性的使用频率,进而转换成各种属性在标注中的权值来提高语义标注的准确性. 武汉大学何克清等人<sup>[18]</sup>提出一种 KMP 算法应用于 Web 服务语义标注中,对 Web 服务进行语义标注是指基于领域本体对 Web 服务进行语义描述,使其能够被计算机所理解,因此标注算法的优劣直接影响到 Web 服务标注的准确率. 其中 KMP 算法的应用使得整个 Web 服务标注过程的准确率和效率较传统语义标注技术有了明显的提高. 上述方法以服务为核心,试图通过提升服务的语义描述能力来增加服务发现的准确率,但以服务提供者为主体的服务描述存在如下问题:(1)对服务属性类型的描述有限,导致服务的查全率不高;(2)服务提供者具有的知识背景出发,其封装的服务功能与非功能并不能体现大众用户的意图,导致服务查准率不高.

本文提出的基于社会标注的服务语义描述方法,指导服务描述的社会标注技术将为服务提供商提供全面、可操作、为大众用户认可的服务封装方法,根据提供商技术能力以及对服务资源认识的不同,从不同层次、不同角度描述服务. 基于多维度的服务语义社会标注方法具有如下优势:(1)可以利用群体支持,解决服务语义标识能力不足的问题;(2)可以由用户对服务的正常运行进行监测,从而降低服务开发和维护的成本;(3)通过标注模板和过滤算法提高标签的有效性. 本文提供的关键方法和技术能帮助服务消费者(用户)从 Web 服务的海洋中准确定位所需服务.

## 6 结束语

Web 服务语义标注是当前语义 Web 服务应用的一个关键性问题. 随着当前网络的发展以及网络资源的开放、共享等特点,Web 用户可以根据自己的需求在网络中查找相应的服务资源满足自己的需要. 但由于用户对资源的使用目的、认识程度的差异

性,导致找到的服务并不一定能够准确满足他们的需求.

本文提出了一种基于社会标注的 Web 服务语义自动浮现方法. 在服务语义社会标注模型的指导下,对服务属性进行维度分类,通过从推荐标签集、候选标签集和自由标签集中选取标签标注服务,使得服务能够浮出语义,呈现大众用户认可的语义资源. 最后通过实验将方法应用于天气领域的 Web 服务语义标注,证明了该方法的实用性和有效性.

本文的主要创新点包括:

(1)从 Web 服务的社会性的角度考虑,支持大众用户对服务进行社会标注,挖掘出普遍认同的真实语义,变主观为统计意义上的客观,提高服务查准率. 与传统的 Web 服务的企业性相比,更能体现大众用户的意图,更适合大众用户准确发现服务.

(2)使用服务语义社会标注体系,可以让每个用户根据自己的关注点的不同去选择某个维度进行标注,也能够根据关注点的不同进行服务查询.

(3)通过对服务属性的度量,使得服务的每个属性类型具有一定的优先级,根据优先级以及标签的出现频率,实现标签的自动排序.

下一步工作的重心是完善该方法,进一步提高服务语义社会标注的准确性,通过更大规模地试用验证方法的正确性和有效性,并在服务语义社会标注的基础上研究服务发现.

**致 谢** 在此,我们向对本文的工作给予支持和建议的同行,尤其是武汉大学软件工程国家重点实验室的彭蓉教授、冯在文博士领导的讨论班上的同学和老师表示感谢!

## 参 考 文 献

- [1] Liang Bang-Yong, Li Juan-Zi, Wang Ke-Hong. Web page recommendation model for the semantic Web. *Journal of Tsinghua University (Science & Technology)*, 2004, 44(9): 1272-1276(in Chinese)  
(梁邦勇, 李涓子, 王克宏. 基于语义 Web 的网页推荐模型. *清华大学学报(自然科学版)*, 2004, 44(9): 1272-1276)
- [2] Verma K, Sivashanmugam K, Sheth A et al. METEOR-S WSDI: A scalable infrastructure of registries for semantic publication and discovery of Web services. *Journal of Information Technology and Management, Special Issue on Universal Global Integration*, 2005, 6(1): 17-39
- [3] Hess A, Kushmerick N. Learning to attach semantic metadata to Web services//*Proceedings of the 2nd International*

- Semantic Web Conference. Florida, USA; Springer, 2003; 258-273
- [4] Patil A, Oundhakar S, Sheth A et al. METEOR-S Web service annotation framework//Proceedings of the 13th International Conference on the World Wide Web (WWW2004). New York, USA; ACM Press, 2004; 553-562
- [5] Ning Da, Rong Peng. A requirements recommendation method based on service description//Proceedings of the 1st International Conference on Cloud Computing (CloudCom2009). Lecture Notes in Computer Science 5931. Springer-Verlag, 2009; 656-661
- [6] Peng Hui, Chen Li-Min, Chang Liang, Shi Zhong-Zhi. Semantic Web service matching based on dynamic description logic. Journal of Computer Research and Development, 2008, 45(12): 2102-2109(in Chinese)  
(彭晖, 陈立民, 常亮, 史忠植. 基于动态描述逻辑的语义 Web 服务匹配研究. 计算机研究与发展, 2008, 45(12): 2102-2109)
- [7] Feng Zai-Wen, He Ke-Qing, Li Bing et al. A method for semantic Web service discovery based on context inference. Chinese Journal of Computers, 2008, 31(8): 1354-1363(in Chinese)  
(冯在文, 何克清, 李兵. 一种基于情境推理的语义 Web 服务发现方法. 计算机学报, 2008, 31(8): 1354-1363)
- [8] Hausmann J H, Heckel R, Lohmann M. Model-based discovery of Web services//Proceedings of the IEEE International Conference on Web Services(ICWS 2004). San Diego, CA, USA, 2004; 324-331
- [9] Kritikos Kyriakos, Plexousakis Dimitris. Requirements for QoS-based Web service description and discovery. IEEE Transactions on Services Computing, 2009, 2(4): 320-337
- [10] Ning Da, Peng Rong. An on-demand model selection method based on O-RGPS domain modeling framework. Journal of Wuhan University (Natural Science Edition), 2009, 55(6): 689-695(in Chinese)  
(宁达, 彭蓉. 一种基于 O-RGPS 领域建模框架的按需模型选择方法. 武汉大学学报(理学版), 2009, 55(6): 689-695)
- [11] Su Zhong, Zhang Xiao-Xun, Zhang Li. Social tagging in Web2.0. Communications of CCF, 2008/10, 2008; 22-29(in Chinese)  
(苏中, 张小洵, 张俐. Web2.0 下的社会标注. Communications of CCF, 2008; 22-29)
- [12] Vojnovic Milan, Cruise James, Gunawardena Dinan. Ranking and suggesting popular items. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(8): 1133-1146
- [13] Yang Kun, Shi Zhongzhi. Exploring Web resources in social annotation services in hierarchy//Proceedings of the 8th International Conference on Grid and Cooperative Computing (GCC2009). Washington DC, USA, 2009; 182-187
- [14] Bao Shenghua, Wu Xiaoyuan, Fei Ben. Optimizing Web search using social annotations//Proceedings of the 16th International Conference on World Wide Web, SESSION: Search quality and precision. New York, USA, 2007; 501-510
- [15] Wu Chao. Analysis of tags as a social network//Proceedings of the 2008 International Conference on Computer Science and Software Engineering (CSSE2008). Wuhan, China, 2008; 651-654
- [16] Xu Bin, Li Juan-Zi, Wang Ke-Hong. Web service semantic annotation. Journal of Tsinghua University (Science & Technology), 2006, 46(10): 1784-1787(in Chinese)  
(许斌, 李涓子, 王克宏. Web 服务语义标注方法. 清华大学学报(自然科学版), 2006, 46(10): 1784-1787)
- [17] Zhang Da-Lu, Lv Tao. Concept frequency-based Web service annotation. Journal of Tongji University (Natural Science), 2008, 36(1): 1395-1400(in Chinese)  
(张大陆, 吕韬. 基于概念频率的 Web 服务语义标注. 同济大学学报(自然科学版), 2008, 36(1): 1395-1400)
- [18] Zeng Cheng, Li Bing, He Ke-Qing. Application of KMP algorithm in Web service annotation. Microelectronics & Computer, 2010, 27(8): 1-4(in Chinese)  
(曾诚, 李兵, 何克清. KMP 算法在 Web 服务语义标注中的应用. 微电子学与计算机, 2010, 27(8): 1-4)



**NING Da**, born in 1984, Ph. D. candidate. His current research interests focus on requirement engineering and service computing.

**HE Ke-Qing**, born in 1947, Ph. D., professor, Ph. D. supervisor. His main research interests include foundation software engineering, software infrastructure, software technology standard.

**PENG Rong**, born in 1975, Ph. D., professor, Ph. D. supervisor. Her current research interests include requirement engineering, service computing, information security.

**FENG Zai-Wen**, born in 1980, Ph. D., lecturer. His current research interests include service-oriented computing, requirements evolution modeling.

**LIU Jian-Xiao**, born in 1984, Ph. D. candidate. His current research interests focus on service computing.

**LI Zheng**, born in 1984, Ph. D. candidate. Her current research interests focus on requirement engineering.

## Background

Service-oriented software development has become more popular application development method. Cloud computing as a new service computing paradigm has become academic and industrial research focus. In cloud computing environment, users are able to use the terminal to find service resources to meet their needs in the cloud service center. However, how to achieve the user's on-demand services has become the bottleneck in cloud computing research.

Semantic Web Service (SWS) provides a solution for aforementioned problem. It is considered from three aspects: first, adding semantic information into the existing Web service description (such as SAWSDL); second, defining directly series of ontology that describe Web service (such as WSMO, OWL-S); third, establishing the corresponding association relation between WSDL elements and ontologies (such as METEOR-S). However, the above Semantic Web service descriptions, mainly from the perspective of service providers to describe services, do not meet the user-centric on-demand services requirements in cloud computing environment.

Social tagging as an activity users involve in has become popular approach of recognition and management of resources in Web2.0 era. For the current social tagging methods related to network resources, the industry has made some efforts. In recent years, there have arisen some social tagging sites, such as dil.icio.us (URLs), YouTube (Videos), flickr

(Photos, Images), Last.fm (Music, Songs), seekda (Web services), existing in some common features that is using the tags to annotate the resources, but with different concerns and do not make good use of tags, resulting in relatively low resource utilization. This research is on a basis of sub-project "requirements meta-model, requirements elicitation and analysis" in the National 973 Program "requirements engineering—the basic research for software engineering of complex systems", with the focus on Web service description and service discovery. The research group has presented a service-oriented requirement elicitation and analysis platform called SKLSEWiki. The major aim of this paper is to improve the description capability of semantic Web service, lay a sound theoretical basis for the further research on accurate service discovery.

This research project was supported by the National Basic Research Program (973 Program) of China under Grant No. 2007CB310801, the National Natural Science Foundation of China under Grant No. 60970017, the Outstanding Youth Foundation of Hubei Province under Grant No. 2009CDA148, the Youth Chenguang Science Project of Wuhan under Grant No. 200950431189, Specialized Research Fund for the Doctoral Program of Higher Education under Grant No. 20090141120020, and the Fundamental Research Funds for the Central Universities under Grant Nos. 6082008, 3101032, 20102110101000119.