

# 智能互联网<sup>\*</sup>)

史忠植 董明楷 蒋运承 盛秋戡 张海俊

(中国科学院计算技术研究所智能信息处理重点实验室 北京 100080)

## Intelligent Web

SHI Zhong-Zhi DONG Ming-Kai JIANG Yun-Cheng SHENG Qiu-Jian ZHANG Hai-Jun

(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China)

**Abstract** Intelligent Web functions essentially as an autonomic entity. The Web automatically regulates the functions and cooperation of related Web sites and available application services. In the paper we will discuss the key issues related to the intelligent Web, such as semantic representation, Agent, resource management, service management, Web mining and so on.

**Keywords** Intelligent Web, Semantic representation, Agent, Web mining, Resource management, Service management

## 1 引言

互联网最早出现在 20 世纪 60 年代,但是直到 90 年代中期才真正引起公众注意并迅速普及。1994 年,全世界仅有 1300 万人使用互联网,但今天这一数字已超过 3.5 亿。互联网从诞生到现在短短几十年的时间,其爆炸式的技术发展速度远远超过了人类历史上任何一次技术革命。然而,从长远发展趋势来看,现在的互联网尚处于发展的初级阶段,互联网技术存在很大的发展空间和发展潜力。

1996 年美国克林顿政府制定了下一代互联网 NGI(Next Generation Internet)计划,与目前使用的互联网相比,它的传输速度更快、规模更大、更安全。世界上大部分国家包括中国在内的第一代互联网建设沿用的都是美国的标准,包括底层技术和核心设备;互联网的巨大变革力量使得各国在下一代互联网研究上展开了竞赛。

网格计算是伴随着互联网而迅速发展起来的、专门针对复杂科学计算的新型计算模式。这种计算模式是利用互联网把分散在不同地理位置的计算机组织成一个“虚拟超级计算机”,其中每一台参与计算的计算机就是一个“节点”,而整个计算是由成千上万个“节点”组成的“一张网格”,所以这种计算方式叫网格计算。这样组织起来的“虚拟超级计算机”有两个优势,一个是数据处理能力超强;另一个是能充分利用网上的闲置处理能力。

互联网的基本特点是它的全球化,超级链接可以提供“任意到任意”的连接,这样就使得世界上所有的计算机可以连接到一起,让用户共享这一无限的资源。然而当今的互联网毕竟还是专门为人们所设计的,网页上文本、图形、图像等各种媒体的涌现都是供人们阅读的,而网络和计算机只是作为一种简单的中介工具。然而对于互联网上庞杂无限的资源,信息过载对于用户来说是最大的问题,如何有效地将各种信息、数

据、知识等资源进行描述、组织和管理,让计算机充分利用这些资源为用户服务。因而我们需要考虑如何利用计算机来阅读、理解和处理这些庞大的资源,并为用户提取出有效的信息和知识,为用户提供各种便利的服务。

智能互联网是下一代互联网的另一种模式,就是有效地将智能融入到互联网中,让具有智能的计算机程序在互联网这种动态开放的无限网络环境中运作,并为人们提供智能的服务。W3C 主席 Berners-Lee 等在《Scientific American》杂志上发表语义互联网文章<sup>[1]</sup>。钟宁、刘际明、姚一豫等人在 IEEE 的《Computer》杂志上组织了一期专集,讨论互联网智能<sup>[2]</sup>。智能互联网本质上是一种自治的实体,它具有自动调整功能,网络中的各个网站之间协同工作,以便进行有效的应用服务。智能主体则是智能互联网中的生灵,它是一种智能的软件实体,能够在智能互联网中自由遨游,为用户提供各种智能的服务。

从智能互联网的总体目标来看,它必须解决以下一些关键问题:(1)首先是语义表达问题,即如何使得网络中的各种信息、数据等资源能够有效地表达并被理解,使得它们成为计算机所具有的“知识”,进而能够被计算机所共享和处理。(2)如何构建各种智能的软件实体,使得它们能够充当智能互联网中的不同角色,具有智能性、自治性和适应性,能够有效地为用户服务。(3)如何有效地将智能互联网中的各种资源组织管理起来,使得它们能够被合理地分配、共享和利用。(4)对于智能互联网中提供的各种服务,如何精确地描述,并将它们有效地组织管理起来,以方便服务的提供和调用。(5)如何在智能互联网中发现和抽取有用的知识并加以利用,以改善智能互联网的结构和利用效率。本文将针对以上问题分别进行讨论。

## 2 语义表达

智能互联网的主要目的就是扩展当前的互联网,使得网

<sup>\*</sup> 基金项目:国家自然科学基金(No. 90104021, 60073019, 60173017);北京市重点自然科学基金(No. 4011003)。史忠植 研究员,博士生导师,主要研究方向为人工智能、机器学习和分布式人工智能等。董明楷 博士研究生,主要研究领域为智能主体、知识表示与推理、网络信息检索、Web 智能。蒋运承 博士研究生,主要研究领域为智能主体和多主体系统、智能互联网。盛秋戡 博士研究生,主要研究兴趣有:多主体系统、面向主体的软件工程、语义网。张海俊 博士研究生,主要研究方向为智能主体、多主体系统、Web 智能。

络中所有信息都是具有语义的,是计算机能够理解和处理的,便于人和计算机之间的交互与合作。要让互联网具有智能,就是要使得它像具有智能的人类社会一样,所有的信息资源都能够被智能的计算机软件实体所识别和处理,并且这些软件实体能够自治地、主动地在网络中漫游,具有很好的适应性和可靠性,能够代替用户完成各种工作,并为用户提供丰富的服务。

因而在智能互联网中,语义表达是首要的问题,也是最为复杂的问题。在人工智能的研究中,知识表示和推理是最基本的问题,也是最为重要的研究内容之一。而知识表示和推理所要解决的,就是如何使得知识是具有语义的,是便于计算机所处理的。在知识的语义表示和推理中,有两个层次的研究:一是在理论上,基于逻辑形式化的研究工作,即以一阶谓词逻辑为代表的各种逻辑体系;二是基于这些逻辑理论的本体论的研究,可以看作是具体的应用研究。下面将分别加以阐述。

### 2.1 知识表示

在知识表示的形式化理论方面,一阶谓词逻辑可以说是最为广泛的工具,它具有很强的表达能力,且具有完备的形式化公理体系,因而一直受到人们的青睐。然而正是由于其很强的表达能力,一阶谓词逻辑是不可判定的,而只是半可判定的,因而不能获得可判定的和有效的推理服务,同时如果不加任何约束,则知识的结构将被破坏。

描述逻辑<sup>[3]</sup>是一种基于对象的知识表示的形式化,也叫概念表示语言或术语逻辑。它建立在概念(Concept)和关系(Role)之上,其中概念解释为对象的集合,关系解释为对象之间的二元关系。描述逻辑是一阶逻辑的一个可判定的子集,具有合适定义的语义,并且具有很强的表达能力。它有清晰的模型-理论机制,很适合于通过概念分类学来表示应用领域,并且提供了许多很有用的推理服务。它可以被认为是从基于框架的表示形式化向着精确的语义特征方向发展。此外,描述逻辑将分类学中表示和推理(术语推理)与在分类学中项的事实或实例的表示和推理(断言推理)区别开来。

描述逻辑主要是用来进行静态知识的表示和推理的。动态描述逻辑<sup>[4]</sup>则是对已有描述逻辑和动作理论的一个综合与扩展,其主要目的就是针对动态世界变化的基本特点,用来对智能互联网中的静态知识和动态知识进行表示和推理。一方面它对网络中各种类型的基本知识进行语义描述,并在此基础上进行推理;另一方面,它还能够对动态概念、动态关系、动作与服务等进行刻画,描述出它们的基本规律与因果关系,进而对智能互联网中的各种活动进行语义表示和推理。当然,知识表示是个很好的思路,很多事实已证明了这一点,但知识表示这门技术,尚未能改变我们的世界。知识表示蕴藏着重要的应用,但要全面挖掘这门技术的潜能,它必须被链入到一个全球性的统一系统中。而智能互联网将提供一种能表达数据和推理规则的统一语言,该语言能让现有知识表示系统中的推理规则用于整个互联网中,从而将逻辑加入到互联网中来。

XML 和 RDF 是两种可用于开发智能互联网的重要技术。利用 XML 可以自己定义标记,例如<zip\_code>,<post\_address>,对网页中的内容进行注释。XML 使用户能将任意的结构添加到文档中,而无需说明结构的具体含义。RDF 通过一些三元组(主题,动词,基本句子的对象)来表达意思。这些三元组可以用 XML 标记表示。在 RDF 中,文档用一系列属性及相应的值来描述特定的事物(如人、网页等任何其他事物)。事实证明,这是种描述大量数据的最自然的方法。主题和

对象均由统一资源标识符 URI 来标识,URI 也被用于表示网页中的链接(URL 是最普通的 URI)。动词也可用 URI 标识,这样能让每个人都可以定义新的概念和新的动词以指向互联网中的某个网页。

### 2.2 本体论

当语义表达通过一定的形式添加到网络资源上之后,下一步工作就是如何使得这些资源被理解和共享。对于互联网上不同的数据资源,它们对同一个概念可能采用不同的标识符,比如对于“邮编”这个概念,有的使用<PostCode>,有的则使用<post\_code>,如何才能让计算机知道它们是标识同一个概念呢?这就需要通过另外一种共有的东西来完成,即采用本体论的方法。

在哲学上,本体是关于存在的本质,也就是什么类型的事物才存在,本体论即是研究关于存在理论的一门学科。人工智能和互联网研究者们同样地把本体作为他们的行话。对他们来说,一个本体就是一个文档或者文件,在该本体中正式地定义了这些术语之间的关系。

在互联网中最典型的本体论都有一个分类法和一组相应的推理规则。分类定义了对象的类和它们之间的关系,而推理规则为互联网提供了一定的推理能力。本体语言就是一种用来描述领域的结构的语言,即对领域本体进行刻画。这些结构可以采用一定的概念(类)和属性(关系)来描述,而本体就由这些结构所表示的公理所组成。由于描述逻辑在语义、可判定性及面向对象的分类表示等方面的优点,因此一般的本体语言可以建立在描述逻辑的基础之上。由 DARPA 组织提出的 DAML+OIL 就是基于描述逻辑而构建的一种本体语言。

对于一个本体语言来说,要求它能够提供有效的推理服务,并且保证推理是可判定的。在本体的设计、集成和使用中,推理都是至关重要的。首先在本体设计阶段,类的一致性以及类之间所隐含的关系都是需要推理的,尤其是对于比较庞大的本体系统。在本体集成中,需要判定本体内部所存在的关系,并把它们整合为一致的层次关系。在本体的使用过程中,需要判断事实集合是否与本体描述(公理系统)一致,同时也可能需要断定某个对象是否是属于某个概念,某些对象之间是否满足一定的关系等。

本体所提供的结构和语义使得一个企业更容易提供一定的服务并且能够使得它的应用完全透明。本体有很多的方法可以增加互联网的功能,同时也使得互联网上的资源是计算机可理解和处理的。它们可以用一种简单的方式来增加网络搜索的准确性。搜索程序可以仅仅搜寻那些指向精确概念的网页而不是搜索所有用模糊关键字的网页。也可以用本体来关联一个网页上的信息和相关知识结构及推理规则,即在网页中嵌入本体,而网页上的术语或者 XML 标识符的意义可以通过指向一个本体指示器来定义。

## 3 主体

主体(Agent)提供了一种很好的问题求解风范,体现在基于主体的系统具有用传统方法难以获得的许多优点<sup>[5]</sup>。主体提供了一种远程智能程序设计的方法。多主体系统放松了对集中式、规划、顺序控制的限制,提供了分散控制、应急和并行处理。并且,多主体系统可以降低软件或硬件的费用,提供更快速的问题求解。

主体与传统的软件系统在很多重要方面是不同的:首先,一个主体具有它所在领域的大量的知识,以及它自己的能力

和目标;其次,主体能从新的信息中不断地扩展它的知识,这些新的信息可以通过观察、与其它主体的通信以及对自己行为的意识而获得;同时,由于这些知识在主体的存储器中不能被清楚地表示,这就表明主体应该能进行推理,能抽取知识;最后,主体要能运用其知识和能力来推理,进行理性的规划,并执行相应的动作。

智能主体对应于多主体系统就如人对应于人类社会。人及人类社会在客观世界中生存、活动并发展,那么智能主体和多主体系统又在什么世界中活动呢?显然,智能互联网就是智能主体的活动环境。首先,智能互联网提供了各种类型的资源,包括各类信息、知识、服务等,这些资源是智能主体能够理解、访问和处理的;同时智能主体也属于智能互联网中必不可少的部分,是智能互联网中最为活跃的元素。在智能互联网中,智能主体扮演着重要的角色,它一方面能为用户提供各种服务,代替用户完成各种类型的信息获取、检索、查询等,并获得互联网上的各种服务,最终满足用户的需求。同时智能主体也能为其它智能主体提供服务,与它们进行交互与协作。

例如,假定你准备在元旦的三天假期中去海南旅游,具体的旅游计划可以直接使用你的个人主体来替你完成。首先你给出关于旅游的一些基本约束,如时间、地点、方式等,然后就可以提交给你的个人主体去制定相应的旅游计划。你的智能主体接受任务后,它依据你的约束条件到智能互联网上去搜索跟海南旅游相关的信息与资源。比如,它首先到各旅游网站上去检索与海南旅游相关的信息,然后把多个不同的信息与用户提交的约束条件进行对照比较,选择合适的信息并制定出相应的供选择的计划。如果某些旅游网站上的旅游服务也是由某些智能主体来完成的,则你的个人主体就可以直接与旅游网站主体进行交互与协作,进一步确定出一些有效的旅游计划。当然,如果你不喜欢参加旅游团,那么个人主体还可以为你再办理其他相关事务,包括网上订购往返机票、预定旅馆等。其中,在订购机票时,个人主体可以根据你的需求而向你推荐最合适的航班以及打折机票。当这些初步计划完成后,个人主体将提交给你供选择。如果你对某个计划比较满意并予以确认后,此时主体就可以执行相应的计划了,而所有这些计划都可以在智能互联网上完成。

通过上述这个例子我们可以看出,主体在智能互联网中具有多种角色,它们既是服务的提供者,也是服务的享受者。对于智能互联网上的各种资源,它都能够理解、访问和处理,即它们能够自动地从智能互联网上获取知识。对于网络上其他主体所提供的服务,主体能够与那些主体进行交互,通过一定的谈判、协作等过程而获得相应的服务,也可以通过委托、授权、分解、组合等方式让多个主体来共同完成一定的服务请求。

在完成这些过程时,有两个基本的条件需要保证。首先,所有这些资源,包括信息、知识与服务等,都应该能够用统一的方式进行表示,并且都基于一定的本体来构建。这样主体就能够以本体的基础来理解、访问和处理这些资源,同时主体之间也可以通过交换本体来达到共识,也增强了主体各自的推理能力。这即是前面所介绍的语义表达部分所要解决的问题。其次,在智能互联网这个社会,需要有一定的法规来约束主体的行为,对不同的主体都有不同的权限约束,以保证智能互联网的秩序。这些约束就可以采用数字签名的方法来解决,即每个不同的主体都有一定的授权,它只能在其授权的范围内履行职责,同时也保证不被其他非法的行为所侵犯。数字

签名是加密的数据块,计算机和主体能够用数字签名来保证信息是由特定的可信赖的用户签发的。在得到消息源的确认之前,主体是不会轻易相信任何它在智能互联网上看到的東西。

由中国科学院计算技术研究所智能信息处理重点实验室研制开发的多主体环境 MAGE<sup>[6]</sup>是一种面向主体的软件开发、集成和运行环境,并且有一套完整的面向主体的方法学。它为网络计算环境下多主体系统的快速开发提供平台,能够工作在动态、开放、异构和分布式等网络环境下,将各种类型的软件系统集成在一起,为用户提供一种智能的网络服务平台。它提供了多种软件重用模式,可以方便地重用以前不同语言编写的主体或非主体软件;它提供了面向主体的软件开发模式,以主体为基本单元,通过封装和自动化主体一般性质,程序员可以通过特殊行为的添加方便地实现自己的应用;这样,通过构建新的软件以及重用旧的软件,应用程序员可以方便地进行各种应用集成;同时它还提供了一个主体之间进行交互通信、协调、合作等活动的网络环境。

#### 4 资源管理

对于未来的智能互联网而言,资源管理问题主要包括以下5个方面<sup>[7]</sup>:节点自治性、异构性、策略的可扩展性、联合分配以及在线控制问题。

其中,自治性是指资源属于不同的组织机构,因此,我们不可能制定一套广泛接受的资源调度和安全等机制。由于资源节点的自治性,因此,各节点本地可能采用异构的资源管理系统。可扩展性需求是指资源管理策略必须考虑具体领域资源的管理结构的多变性,而不影响其他节点的配置。多个应用同时请求资源是联合分配问题的由来,节点的自治性以及资源分配过程中可能的失败都需要我们提供专门的资源分配,使用监控等机制。由于资源需求和资源自身可能随时改变,因此,资源提供者和资源需求者之间需要进行在线协商和控制。图1描述了智能互联网中的资源管理架构。

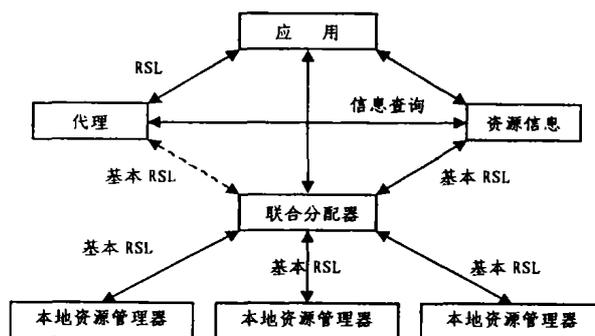


图1 资源管理架构

资源管理器用于解决节点自治性和异构问题,它为本地资源管理工具、安全机制等提供了良好的接口。在线控制和策略的可扩展性是由支持协商的可扩展资源规约语言(Resource Specification Language, RSL)实现的。而资源代理负责从高层应用请求向各资源管理器请求的映射。联合分配通过在资源联合分配器中定义相应的资源分配策略实现。

应用与资源代理、资源联合分配器、资源管理器之间通过RSL进行通信,RSL表示具体的资源信息,如资源需求等。资源代理负责将高层的RSL内容转换成更具体的基本资源请求。基本资源需求包含了资源的完整地址。联合分配器接收基本的资源请求后,负责对相关的资源节点进行协调和管理。信

息服务模块负责提供有效的资源信息,用于资源定位,确定相关的资源管理器,确定资源的属性以及提供从高层资源规格说明转换成对具体的资源管理器的请求过程中涉及的其他信息。

## 5 服务管理

服务管理是智能互联网的重要研究课题,目前的 Web 只能为用户提供简单的静态信息,而在智能互联网上除了能提供静态信息外,还必须能提供动态的服务,从而必须要对智能互联网上的服务进行有效的管理。服务管理是一个很复杂的问题,具体来讲,服务管理主要包括以下内容:服务描述、服务匹配、服务组织、服务调用、服务协商、服务组合以及服务监控等<sup>[9]</sup>。下面对它们进行简单的介绍。

服务描述就是如何表示服务以利于计算机理解,它的任务是实现一种服务描述语言来描述服务<sup>[10]</sup>。服务描述语言要具有以下一些主要特点:表达能力强,不仅能表达信息、知识,还要能表达服务;灵活性,能适应多种情形;推理功能,能自主地完成服务匹配、服务调用、服务组合及服务验证等一系列服务管理问题;基于语义的描述,因为智能互联网是一个开放、异构和动态的环境,各用户描述服务时所用术语不统一,如果只考虑在语法层上描述服务(如基于关键词的描述),就不能解决系统间的互操作问题,在描述服务时,应该在语义层上描述;支持服务的协商机制,因为服务的组合离不开服务提供者与服务请求者之间的协商;支持数据类型,服务的输入和输出等属性不是简单的字符串,应该要考虑数据类型的使用;要考虑服务质量与服务效率的关系,因为这两者是一对矛盾,有的用户以服务质量为优先条件,有的用户以服务效率为优先条件,而有的用户要考虑两者的平衡。

服务匹配即帮助用户定位他们所需要的服务,它的大致过程如下:(1)服务提供者向服务中介注册它们的服务;(2)服务中介存贮这些服务提供者的服务,并动态地形成全局术语本体;(3)服务请求者向服务中介请求是否有提供它所需服务的提供者;(4)服务中介采用服务匹配算法将服务请求者的服务请求与它所存贮的服务提供者的服务进行匹配,并返回结果。看起来服务匹配过程似乎很简单,实际上该过程非常复杂,因为智能互联网是一个开放、异构和动态的环境,它上面的信息、资源和服务的位置和属性等经常发生变化,不同团体或组织所使用的服务描述术语不统一等。由此可以看出,服务匹配主要解决两个问题:首先是服务匹配的交互协议以及通信语言。多主体系统即是一个简单的智能互联网,就多主体系统中关于服务匹配的交互协议,目前一般采用匹配主体结构和代理主体结构;关于通信语言,目前一般采用 KQML 或 FIPA-ACL。其次是服务匹配算法,目前主要有 3 种:近似匹配算法、精确匹配算法和插入匹配算法。

服务组织就是服务的存放问题,因为智能互联网是一个庞大的开放系统,向服务中介注册服务的提供者数量是巨大的,如何对这些注册的服务进行合理有效的组织存放是一个重要的问题。在此基础上实现服务高效的服务查找过程,提高服务匹配的速率。

服务调用就是具体使用服务提供者的服务,即按照服务提供者的要求,传给它必要的入口参数等,等待它的返回结果。该过程的主要问题有:找到服务提供者的位置以及服务参数(或服务结果)的传输协议等。

服务协商即是服务提供者与服务请求者之间的协商,主

要有 3 个问题:使用什么协商协议?协商的问题(内容)是什么?以及协商者使用什么样的推理模型?对于一个服务请求者,如果有多个服务提供者能满足它的要求,这就存在服务协商问题,协商的内容包括服务时间、价格、服务质量等,即服务提供者通过协商来选择它“最满意”的服务提供者。

服务组合是服务管理的一个关键问题。简单地说,它要解决的问题是:当服务请求者提出服务请求时,系统中没有一个单独的服务提供者能满足该服务请求者的要求,但是,通过几个服务提供者的“合理组合”可以满足服务请求者的要求,即几个服务提供者联合起来可以为服务请求者提供它所需的服务请求。它要解决的关键问题主要有:任务分解、原子过程描述、服务组合描述、过程控制信息的管理以及结果的组合问题等。目前实现服务组合的方法主要有:工作流方法、状态图方法、过程模式方法以及过程本体方法,这些方法的共同点是服务过程用一种方法表示出来,根据服务过程来实现服务的组合。也有的研究工作根据组织角色来实现服务组合。

服务监控就是用户可以随时跟踪、监督服务的执行情况,它主要解决如下问题:用户随时查询服务的执行情况、查询完成服务的全过程或者某些子过程、查询服务失败的原因以及服务结果的安全性和可靠性等。对于这方面的问题,目前还没有一个具体的解决方案,有待深入的研究。

智能互联网上的服务管理是一项重要而又艰巨的任务,它的主要目的就是为用户提供高效的服务。要解决智能互联网上的服务管理,涉及多门学科:人工智能、软件工程、本体论、逻辑、网络、博弈论以及决策论等。

## 6 互联网挖掘

人类从工业社会向知识社会演进时,政治经济中心正从“生产”转向“发现、发明和创新”。知识正在成为创新的核心,知识创新成为知识经济发展的最主要的动力源泉。知识经济对物质文明发展能够发挥巨大的推动作用,依靠无形资产的投入来实现可持续发展的,推动经济全球化发展<sup>[7]</sup>。

互联网挖掘是从互联网中发现和抽取有用的知识并加以利用,以改善互联网信息的利用效率和效果。互联网挖掘主要包括互联网内容挖掘、结构挖掘、文档自动分类与聚类、动态信息挖掘、事件自动生成、用户信息挖掘等。

互联网内容挖掘是从网页的内容信息中抽取知识,又分为文本文档和多媒体文档的挖掘。互联网页面的复杂程度远远高于普通的文本文档,它是以某种格式(HTML 或 XML)呈现的半结构化数据,数据结构隐含,模式信息量大,模式变化快。目前的互联网内容挖掘研究主要集中在基于文本文档内容的检索、信息过滤和提炼、重复数据消除、数据模式抽取、中间形式表示、文档资料聚类和分类、文档资料结构提取。

互联网结构挖掘将从互联网文档的结构信息中推导知识,所利用的信息不限于文档之间的超链结构,还包括文档的内部结构,以及文档 URL 中的目录路径结构等信息。互联网超链结构与期刊的引用不同。互联网网页的作者创建一些超链是为了别的目的,例如导航、广告。为了商业或竞争的目的,不会把互联网网页指向他的竞争权威网页,而权威网页很少提供说明描述。鉴于上述原因,研究网页的权威性和广泛性等结构信息将有利于互联网结构的研究。

互联网文档自动分类是用大量带有类标识的文本,对分类准则或模型参数进行训练,然后用训练得到的结果对未知

(下转第 10 页)

因此,应用 AOP 时仍然要遵从良好的设计原则,只分离需要分离的关注点。

关于贯穿特性的一个重要问题是理解谁“贯穿”了谁。实际上,发生“贯穿”关系的双方是互相贯穿的。分别从商业业务功能关注点的角度和安全、日志等系统关注点的角度,会得到完全相反的“贯穿”结果。由于目前人们对系统进行分解时,多数将商业业务功能作为主导维度,因此,安全、日志等系统关注点就是主要的“贯穿特性”。

AOP 正在迅速发展,其思想已经超越了程序设计阶段,正应用于软件生命周期的不同阶段。AOP 正在许多领域得到应用,比如中间件、安全、容错、服务质量、操作系统等。不过,AOP 目前并没有完全成熟,各种概念和实现机制尚在不断完善中。在实践中,AOP 没有得到工业界的大规模推广,没有经历开发大型系统的实际考验,因此,目前断言 AOP 有效与否尚为之过早。我们需要更深入地研究和实验,回答下述若干问题:AOP 对大型项目有效吗?它最适合解决什么问题?什么构造形式最适于声明“贯穿特性”?这些问题的回答,需要学术界和工业界的广泛参与,需要更多应用的支持。

我们北京大学软件研究所在研究基于软件体系结构的构件组装过程中,将 Aspect 的概念引入到软件体系结构和构件运行平台中。通过在 ABC/ADL 中加入相应的语言元素,在设计阶段描述系统贯穿特性<sup>[10]</sup>。同时,在构件运行平台 PKUAS 上<sup>[14]</sup>,通过构件容器的截取器(Interceptor)机制,提供在运行时刻动态组装编排 Aspect 的机制。PKUAS 中的安全、事务等,都通过这种方式实现。

可以肯定的是,AOP 不会是处理关注点分离问题的最后解决方案。新的更有效的程序设计技术仍将在现有的基础上涌现,从而推动软件工程不断进步。

(上接第4页)

类别的文本进行识别。文档聚类是将文档集合分成若干簇,要求同一簇内的文档内容的相似度尽可能地大,而不同簇间的相似度尽可能地小。互联网动态信息挖掘主要挖掘互联网内容、结构和访问模式的变化。

用户信息挖掘是从用户对互联网的使用方式和行为中挖掘有用的模式,用于对互联网信息的合理组织和服务质量的改进。挖掘的结果通常是用户群体的共同行为和兴趣,以及用户个人的检索偏好、习惯和模式等。通过用户信息挖掘,可以进行主动服务,改善服务质量。

多策略数据挖掘平台 MSMiner<sup>[7]</sup>是由中国科学院计算技术研究所智能信息处理重点实验室研制的,它提供决策树、支持向量机、粗糙集、模糊聚类、基于范例推理、统计方法、神经计算、可视化等多种数据挖掘算法,支持特征抽取、分类、聚类、预测、关联规则发现、统计分析等数据挖掘功能,并支持高层次的决策分析功能。

**结束语** 智能互联网是互联网发展的一种模式,它有效地将智能融入到互联网中,让具有智能的计算机程序在互联网这种动态开放的无限网络环境中运作,共享网络资源,为人们提供高质量的服务。智能互联网将在语义网络的架构下,利用主体、资源管理、服务管理、数据挖掘等技术,开创互联网的新局面,提供智能化的信息基础设施<sup>[11]</sup>。

## 参考文献

- 1 Kiczales G, Lamping J, et al. Aspect-Oriented Programming. In: Proc. of ECOOP'97, 1997
- 2 Ossher H, Tarr P. Using Multi-dimensional Separation of Concerns to (Re)Shape Evolving Software. CACM, 2001, 44(10)
- 3 Bergmans L, Aksit M. Composing Crosscutting Concerns Using Composition Filters. CACM, 2001, 44(10)
- 4 Lieberherr K, Orleans D, Ovlinger J. Aspect-Oriented Programming With Adaptive Methods. CACM, 2001, 44(10)
- 5 Elrad T, Filman R E, Bader A. Aspect-Oriented Programming. CACM, 2001, 44(10)
- 6 Elrad T, Aksit M M, Kiczales G, Lieberherr K, Panelists H O. Discussing Aspects of AOP. CACM, 2001, 44(10)
- 7 Pace J A D, Campo M R. Analyzing The Role of Aspects in Software Design. CACM, 2001, 44(10)
- 8 Kiczales G, et al. Getting Started with AspectJ. CACM, 2001, 44(10)
- 9 Constantinides C A, et al. Designing an Aspect-Oriented Framework in an Object-Oriented Environment. ACM Computing Surveys (CSUR), March 2000
- 10 Mei Hong, et al. ABC/ADL: An ADL Supporting Component Composition. In: Proc. of ICFEM 2002
- 11 Laddad R. I want my AOP! <http://www.javaworld.com/javaworld/jw-01-2002/jw-0118-aspect.html>
- 12 Aspect-Oriented Software Development Web site. <http://aosd.net>
- 13 邵维忠, 杨芙清. 面向对象的系统分析. 北京:清华大学出版社, 1998
- 14 黄翌, 王千祥, 曹东刚, 梅宏. PKUAS: 一种面向领域的构件运行支撑平台. 电子学报, 已录用

## 参考文献

- 1 Berners-Lee T, Orlassila J H. The Semantic Web. Scientific American, May, 2001
- 2 Zhong Ning, Liu Jiming, Yao Yiyu. In Search of the Wisdom Web. Computer, IEEE CS, 2002, 35(11): 27~31
- 3 Baader F, Nutt W. Basic description logic. In: F. Baader et al. eds. The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press, 2002
- 4 Dong Mingkai, Zhang Haijun, Shi Zhongzhi. Dynamic Description for Agent Programming. In: Proc. of CRBCIP 2002. 11~19
- 5 史忠植. 智能主体及其应用. 科学出版社, 2000
- 6 <http://www.intsci.ac.cn/>
- 7 史忠植. 知识发现. 清华大学出版社, 2001
- 8 Czajkowski K, Foster I. A Resource Management Architecture for Metacomputing Systems. In: Proc. IPPS/SPDP '98 Workshop on Job Scheduling Strategies for Parallel Processing, 1998
- 9 The DAML Services Coalition, DAML-S: Web Service Description for the Semantic Web. In: The First Intl. Semantic Web Conf. (ISWC), Sardinia, Italy, June, 2002
- 10 蒋运承, 张海俊, 董明楷, 史忠植. 多主体系统中的动态服务匹配. 已投软件学报, 2002
- 11 史忠植. 高级计算机网络. 电子工业出版社, 2001