

摘要

语义 Web 服务是结合语义 Web 和 Web 服务两方面技术的一个热门的研究课题。目前与语义 Web 服务有关的研究工作大部分针对语义 Web 的层次结构,使用 OWL-S 语言,而本文的研究工作着重于语义 Web 与知识工程的结合。在知识工程中,用本体描述领域知识和问题解决方法(Problem Solving Methods, PSMs),不但使知识工程更加科学化和规范化,而且通过将知识工程与语义 Web 结合,使 PSMs 也纳入了 Web 服务的范畴。

通常基于知识的系统可分为两部分:分析型和综合型。其中,分析型系统一般包括分类、诊断、评估、监测等类型,但目前分析型系统并没有一个通用的处理模型。在课题组前期工作的基础上,本文针对实际领域所具有的知识的多少,对传统的分析型系统实例进行了扩展,归纳总结为偏差问题,并提出了相应的 PSMs。通过抽取 PSMs 的主要概念,以及分析概念之间的关系,给出了偏差问题的语义 Web 服务组合架构—偏差型语义 Web 服务,并通过 WSMO (Web Service Modeling Ontology) 实现了实验性可在网上发布的偏差型语义 Web 服务,同时对语义 Web 服务的自动组合进行了研究。

本文的主要工作和创新如下:

- (1) 根据偏差问题中领域知识的多少,将知识工程中基于知识的分析型系统加以扩展后定义为偏差问题,并给出了此类问题的 PSMs。
- (2) 通过对 PSMs 的概念的分析,给出了与其对应的偏差问题的语义 Web 服务组合架构,呈分层树状结构,将偏差问题转化为一系列的语义 Web 服务的组合。为了使组合架构支持服务执行关系的管理,设计了服务组合的控制算子,为组合服务转为可执行过程提供了基础。
- (3) 基于偏差问题组合架构,在 WSMO 平台上嵌入 OCML (Operational Concept Modeling Language) 语言实现的服务组合控制算子,并构建了实验性的可在网上发布的偏差型语义 Web 服务。
- (4) 根据偏差型服务的特点,提出与或图方式的自动组合方法,包括与或图解图的搜索及解图转换为可执行工作流的方法,并给出了对偏差问题的语义 Web 服务组合架构进行处理的体系结构。
- (5) 语义 Web 服务的知识处理平台是课题组最终的研究目标。本文给出了关于知识处理平台的粗略思想,并对其中的几个关键问题进行了讨论。

关键词 偏差问题; 语义 Web 服务; 服务自动组合; 问题解决方法; WSMO

Abstract

Semantic Web Services is a hot research work, which is composed of Semantic Web and Web Service. Currently research work related to semantic web services is largely originated from semantic web groups, mainly based on OWL-S language, but this paper focuses on the composition research of semantic Web and knowledge engineering. In knowledge engineering, ontology is used to describe the certain field knowledge and problem-solving methods (PSMs) . It is not only promoting the knowledge engineering more scientific and standardized, but also conducting the PSMs into semantic Web scope through the composition.

In generally, knowledge-based system could be divided into two classes: analysis system and integration system. Analysis system generally includes classification, diagnosis, evaluation, monitoring etc, but there is not a common model in analysis system. Based on the even research of seminar and traditional analysis system, the paper expands the scope of analysis system, sum up as the deviation problem according to the quantity of actual knowledge and proposes the corresponding PSMs. Through abstracting the main concept of PSMs and analyze the relationship between concepts, the paper provides the composition framework of Semantic Web services of deviation problem, build an experimental knowledge-based Semantic Web Services which could be published on-line based on WSMO (Web Service Modeling Ontology) , and study about how to achieve the automatic web services composition.

The main works and innovation is as follows:

- (1) According to the quantity of domain knowledge, expand the traditional analytics system of knowledge engineering as deviation problem, and propose the corresponding PSMs.
- (2) Through the concepts analysis of PSMs, proposes the corresponding composition framework of semantic web services of deviation problem, transfers the deviation problem to the composition by a series of semantic web services. The composition framework represents as layered tree structure. To manage the execution control, expanded the combination operator which provide the implement foundation of transferring the web composition to executable process.
- (3) Based on the composition framework, embed the control operator by OCML (Operational Concept Modeling Language) in WSMO platform, build a

- motivation example which would be published online.
- (4) According to the feature of knowledge-based services, study how to achieve the automatic composition of knowledge-based Semantic Web Services, propose an approach to support the automatically composition and convert it into executable workflow by And/Or graph, and provides the architecture which could deal with the composition framework of deviation problem.
 - (5) Knowledge management plat of semantic web services is the ultimate goal of our seminar. This paper proposes rough idea about plat, and discusses three key issues.

Keywords deviation problem; semantic web services; automatic services composition; problem solving method; WSMO

独创性声明

本人声明所提交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京工业大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签名： 贾慧彤 日期： 2009.6.1

关于论文使用授权的说明

本人完全了解北京工业大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

（保密的论文在解密后应遵守此规定）

签名： 贾慧彤 导师签名： 王 鑫 日期： 2009.6.1

第 1 章 绪论

1.1 引言

互联网在当今社会的影响不言而喻，它的基本功能从知识的查询扩展到服务的提供，但是传统的 Web 服务技术缺乏计算机可理解的语义，限制了 Web 服务的自动化发展，于是结合了语义 Web 技术和 Web 服务技术的语义 Web 服务就应运而生。目前与语义 Web 服务有关的研究工作大部分来源于语义 Web 的层次结构，基于 OWL-S 语言，但还有一部分的研究工作来源于语义 Web 和知识工程的结合，通过对基于知识的系统的推理部分-PSMs 的研究，采用本体描述，从而与语义 Web 服务建立对应。此领域的研究更注重与特定领域的结合，代表系统是 IRS (Internet Reasoning Service)，IRS-II^[1]使用基于知识系统的可重用组件构建知识模型，将一个 PSMs 看成是一个 Web 服务，IRS-III^[2]实现了 WSMO 体系^[3]，把重点转向 Web 服务的组合，并主要支持特定的商业领域，提供了一个半自动的 Web 服务的组合方法^[4]。2007 年，IRS-III 对目标 (WSMO 顶层组件)重新提出了新的模型^[5]，设计和使用目标模板作为引导组合操作的抽象模型，应用到了生物工程学^[6]、医学^[7]、空间数据集成^[8]、政府机构^[9]等领域。

语义 Web 服务的自动组合是指 Web 服务在语义的基础上自动进行组合以共同完成一项任务，是关于语义 Web 服务的研究热点与难点问题。目前，国内关于语义 Web 服务的自动组合研究，基本上都是使用 OWL-S，组合技术基本来自人工智能的规划技术。而基于知识工程路线的语义 Web 服务的研究使用 WSMO 体系，与 OWL-S 相比，它覆盖了 OWL-S 的大部分描述元素，并引入了额外元素^[10]。WSMO 在语义层上面区分目标和 Web 服务，因此可以支持功能驱动的服务的匹配和调用。WSMO 的另一个特征是桥组件的使用，所有的数据绑定都使用桥转换过程，相对于把桥和 Web 服务同等对待的 OWL-S、BPEL4WS，增加了可理解性，较好地解决了异构问题^[11]。全自动的语义 Web 组合仍是此路线研究的目标，但目前尚未看到有关的报告。

基于知识的系统可分为两部分：分析型和综合型，其中，分析型系统一般包括分类、诊断、评估、监测等类型^[12]。在以往的文献中，对于分析型系统并没有一个通用的处理所有类型的模型，只是涉及类型之间的一些共同的特征。因此，这种方法不适合于采用 Web 服务自动组合方式来处理基于知识的系统。

为了开展语义 Web 服务的自动组合研究，需构造处理基于知识系统的通用模型。课题组针对实际领域所具有的知识多少，在对传统的分析型系统实例

研究的基础上扩展了研究范畴,将所研究的问题归纳后,称之为偏差问题。并通过抽取偏差问题的主要概念,及分析概念之间的关系,构建偏差问题的 PSMs。在此基础上,进一步将 PSMs 转化为偏差问题的语义 Web 服务组合架构,并在 WSMO 体系下将其实现,同时对语义 Web 服务的自动组合方法进行研究。

1.2 研究背景

1.2.1 语义 Web

传统的 Web 资源由于缺乏明确的描述而无法清楚地表达 Web 资源的语义,如果信息能以语义良定义形式存在,则将大大提高资源的共享能力。目前热门的语义 Web 技术,是指用含有语义的标记语言来描述 Web 服务的语义,使 Web 服务成为计算机可以理解的实体,其基本出发点是:将机器可读的信息加入 Web 页面,采用本体(ontology)的天然共享的技术精确定义 Web 资源中的词汇,利用知识表示技术中逻辑(Logic)的优良的形式化结构和推理规则实现 Web 资源的自动推理。语义 Web 的概念由 Tim Berners-Lee 于 1998 年首次提出。2000 年 Berners-Lee 在 XML 国际会议上提出了语义 Web 的层次结构,它主要基于 XML 和 RDF(Resource Description Framework),并在此之上提供本体和逻辑推理规则。在语义 Web 的发展过程中,世界各地的各个研究机构和科研人员都在密切关注着它的发展,从理论到实践、从标准到规范,都在做着不懈的努力。其中,语义 Web 标记语言的研究是其核心内容。美国 Maryland 大学创立了 SHOE(Simple HTML Ontology Extension),它的语法基于 HTML,语义基于 Horn 逻辑。欧共体的 On-To-Knowledge 项目[Ontok]资助开发了 OIL,它是在 RDFS 层之上增加定义的一层,基于描述逻辑(Description Logic)以提供形式化的语义和推理功能。美国 DARPA 资助的 DAML(DARPA Agent Markup Language)[DAML]项目产生了 DAML,由于 DAML 更多地和 OIL 保持了一致性,因此后来的版本称为 DAML 十 OIL。W3C OWL(Web Ontology Language)工作组在 DAML+OIL 的基础上进行了一定的改进,2004 年 2 月 10 日,OWL 正式成为 W3C 推荐的标准。DAML+OIL 和 OWL 都是建立在 RDF 和 RDFS 之上,基于描述逻辑以提供形式化的语义和推理功能,从而为语义 Web 提供了 ontology 层。OWL-S 来自于 DAML-S,它使用 OWL 作为本体语言从语义上描述 Web 服务^[13]。

1.2.2 知识工程中的问题解决方法

知识工程^[14] (Knowledge Engineering) 最初是由费根鲍姆 (Feigenbaum) 于 1977 年提出来的。当时, 他认为“知识工程是用人工智能的原理和方法, 对那些需要专家知识才能解决的应用难题提供解决的手段”。1980 年本体首先被 John McCarthy 从哲学领域引入计算机领域。1985 年 William J. Clancey^[15] 提出了启发式分类的知识处理思想。1998 年 Perez 等人提出比较完善的本体思想后, 许多本体库被开发出来。知识工程的研究也与本体结合, 主要方向可分为两部分: 一个是建立基于领域本体的本体库; 另一个是关于 PSMs 及其方法论的研究, 如 CommomKADS, 主要研究知识密集型领域的 PSMs。PSMs 描述了独立领域的一个通用的推理行为。这些工作通过对基于知识的系统 (通常是指应用于知识密集型领域的系统) 的推理部分的研究, 采用本体描述, 从而与语义 Web 服务建立对应, 走出了一条基于知识工程的语义 Web 服务的研究路线。1998 年, Benjamins and Fensel 的 Editorial: problem-solving methods 文章使 PSMs^[16] 开始进入一个比较成熟的阶段。

在 Clancey 提出的启发式分类的知识处理思想基础上, 2000 年, Enrico Motta 和 Wenjin Lu 在 IBROW (An Intelligent Brokering Service for Knowledge-Component Reuse on the World-Wide Web) 工程中提出了一种分类的模型化方法, 在他们的研究中将与已知的类别匹配时可能发生的情况根据匹配标准的不同分为四种, 用以处理分类问题^[17]。

知识系统中有一类领域所具有的知识较少 (下述偏差问题的第三种类型), 既没有长期经验的积累, 也没有大量的基础数据可用来学习规律, 因此也就不具有与偏差对应的层次类别, 仅有一些偏差及其产生原因之间简单的对应关系。在以前的工作中, 课题组基于对此领域的一些实际问题的研究, 对 Enrico Motta 和 Wenjin Lu 的方法进行了改进, 采用属性和属性值分离的方法, 重新构造了匹配的标准, 并在试卷分析、超市经营分析、球赛数据分析等领域进行了实例分析^[18]。另外一个同学应用 UPML 框架实现了此类的偏差 PSMs^[19], 包括: 构建了 Jess 基础本体, 完成了偏差任务 PSMs 的任务组件、PSMs 组件、本体组件和桥组件的本体描述及 Jess 实例构建。

1.2.3 语义 Web 服务的知识工程路线

Fensel, Motta 和 Benjamins 继续 PSMs 的研究, 并在 IBROW 项目 (主要目标是开发一组支持万维网重用的组件群, 1999) 的第二次大会 (2nd IBROW meeting, 17-18 May, 2000) 上提出了一个称为 UPML (Uniform Problem-solving

Method description Language 统一问题解决方法语言)的框架^[20],它是一个 PSMs 开发语言,其功能是描述组件、描述连接组件的适配器和这些组件应满足的约束。UPML 定义了四个主要的知识组件:任务,问题求解方法,领域模型和本体。在项目的第三次会议(3rd IBROW meeting, 11-12 Oct. 2000)上提出了 OCML 建模语言,OCML 是建立在 LISP 之上的宏,因此具有 LISP 的天然 AI 优势。OCML (Motta and Lambert,2000)^[21-22]通过一组已经定义类型的数据结构(函数、关系、类、实例和规则)来支持知识模型的构建,同时还包括本体和 PSMs 的定义。目前,在 KMI 的很多项目(包括 IRS 系统)中都使用 OCML 来模型化知识结构,如知识管理、本体开发、电子政务和基于知识的系统开发等。现在已有一些可重用的模型化组件来支持 OCML,它们提供了一些知识组件的可重用性。

IRS 开始于 IBROW 工程,是 KMI (knowledge media institute of open university)组织的一个语义 Web 服务框架,最初由英国的 Open University 大学提出并组织研究,它支持异源地语义 Web 服务的发布、定位、组合和执行,且支持 Web 服务的语义推理规则。目前有两个实现的版本,是 2003 年开始的 IRS-II 和 2004 年开始的 IRS-III。IRS-II 已经应用于 MIAKT 项目,目前 IRS-III 应用于 DIP 项目、Super 项目、Luisa 项目、LHDL 项目和 WSMO 工作小组^[23-24]。

WSMO 工作组是欧洲语义 Web 组织(ESSI)下属的一个团队,是在 SEKT 项目、DIP 项目和 ASG 项目中发展起来的。小组中很多语义领域专家是 Open University 成员,也都曾是 IRS 小组成员,在研究 IRS-II 后继续 WSMO 的研究,IRS-III 版本就已经按照 WSMO 来构建 Web 服务本体。ESSI 的 WSMO 工作组的目的是通过这几个关键的欧洲语义项目更好的发展语义 Web 技术,进一步规范语义 Web 服务的语言标准和构建语义 Web 服务体系结构和平台。在提炼和扩充的基础上提出了 WSMO (web service modeling ontology),即 Web 服务模型本体框架标准,该标准已于 2005 年 6 月提交给 W3C 组织审核^[25-26]。

WSML 工作小组是 WSMO 工作小组的一部份,WSML 工作小组致力于开发一种称为 WSML(Web Service Modeling Language)的语义 Web 本体(WSMO)模型化语言。因此具有以下任务:开发一种适合于语义 Web 服务的模型化语言,提供一个基于规则的语义 Web 服务语言^[27-28]。

WSMO Studio^[29]是 DIP 项目、InfraWebs 项目、SemanticGov 项目、SUPER 项目部份基金的成果。它是一个开源的语义 Web 服务和语义模型本体(WSMO)的语义商业模型环境平台。它由一系列可嵌入 Eclipse 的第三方软件组成。WSMO Studio 现在已支持 Windows、Unix、OS 平台上的使用。它的版本从 2005 年的 0.1.1 版本到目前的 0.7.2 版本。2005 年 6 月推出的 WSMO 建立语义 Web

服务应用 API 和实现参考的 WSMO4J^[30]; 2005 年 11 月推出的关于 WSML 语言的一系列工具, 如 WSML 推理工具 WSML DL Reasoner 和 WSML 语言校验工具 WSML Validator。

1.3 本文的主要工作

在以往的工作基础上, 本文根据实际领域所具有的知识多少, 在对传统的分析型系统实例进行研究的基础上进行了扩展, 归纳总结为偏差问题。通过抽取偏差问题的主要概念, 以及分析出概念之间的关系, 提出了相应的 PSMs。基于 PSMs 提出了偏差问题的语义 Web 服务组合架构, 本文称之为偏差型语义 Web 服务。WSMO 是构建知识系统的语义 Web 服务的平台, 因此选择 WSMO 平台实现了可发布的偏差型语义 Web 服务, 同时对语义 Web 服务的自动组合进行研究。为达到这一目的需要五个方面的工作:

(一) 针对实际领域所具有的知识多少, 对传统的分析型系统及其扩展的实例进行研究, 抽象出偏差问题的主要概念, 以及分析出概念之间的关系, 提出相应的 PSMs。

(二) 基于 PSMs 提出了偏差问题的语义 Web 服务组合架构, 将偏差问题的求解转为语义 Web 服务的组合。偏差问题的语义 Web 服务组合架构由一系列小粒度的语义 Web 服务组合而成, 分为知识级语义 Web 服务和原子级语义 Web 服务两种类型。为了支持服务执行关系在架构中的正确反映和执行控制, 设计了在组合架构中的组合控制算子, 为组合服务转为可执行过程提供了基础。

(三) 偏差问题的语义 Web 服务组合架构的实现与发布。基于 WSMO 平台, 将偏差问题的组合架构转换为语义 Web 服务描述, 实现了偏差型语义 Web 服务, 通过 IRS 进行发布。使用 OCML 语言对服务组合控制算子进行定义, 将其嵌入到 WSMO 系统中, 并加以实现。

(四) 对偏差问题的语义 Web 服务组合架构进一步扩展, 研究架构中服务的自动组合。提出一种使用与或图组合 Web 服务的方法, 使用扩展的启发式算法获得最优解图, 并支持与或解图到可执行工作流的转换, 此方法实现了组合架构的分层树状结构到可执行的线性工作流的转换。

(五) 我们的最终目的是建立一个知识处理平台。由于知识的复杂性, 只给出了语义 Web 服务的知识处理平台的粗略思想和关键几个问题的讨论。

本文的研究将基于知识的系统应用通过提供语义 Web 服务的方式扩展到互联网上, 不仅体现了知识的共享与重用, 而且普及了基于知识的系统的应用。同时, 对 PSMs 采用新的标准开展进一步的分析、挖掘和归纳, 进一步加深了

对其本质的了解。建立基于知识的语义 Web 服务组合架构,使得基于知识的系统应用有一个更巩固的理论和工程基础,从而有助于使知识工程与其它工程看齐。基于 PSMs 的偏差型语义 Web 服务的自动发现与组合算法的研究,对解决语义 Web 服务自动组合这个热点与难点问题提供了一个新思路,有助于此问题的实际解决。

1.4 论文组织结构

本文的整体结构如下:

第一章,首先对偏差问题的研究背景、研究意义、研究现状以及本论文的研究内容进行了概括性的介绍。

第二章,主要介绍偏差问题概念,包括其特点,领域分类等。提出偏差问题的 PSMs,并对其具体执行方法进行了详细的解释。

第三章,主要介绍将偏差问题转换为语义 Web 服务的组合架构。组合架构由知识级语义 Web 服务和原子级语义 Web 服务两种组成,呈分层树状结构。扩展组合架构加入了控制算子,为组合服务成为一个可执行过程提供了执行控制。

第四章,主要介绍偏差型语义 Web 服务的实现,包括 WSMO、IRS 平台的介绍和实例构建一个偏差问题的语义 Web 服务组合,并在 WSMO 体系中加入了 OCML 语言实现的控制算子。

第五章,主要介绍服务自动组合方法,服务自动组合方法采用一种基于与或图的方法,包括基于相似性的任务分解,优化解图求解,解图到执行工作流的转换,和工作流中的交互和正确性检验,并使用实例进行实验说明。

第六章,主要介绍语义 Web 服务的知识处理平台,包括其粗略思想和几个关键问题的讨论:偏差型语义 Web 服务的组合和工作流,知识级语义 Web 服务集合,原子级语义 Web 服务集合及粒度大小,Web 服务语义功能描述。

第2章 知识系统中的偏差问题

基于知识的系统可分为两部分：分析型和综合型。在以往的文献中，对于分析型系统并没有一个通用的处理所有类型的模型，只是涉及类型之间的一些共同特征。因此，这种方式不适合于采用 Web 服务组合方式进行知识系统问题的处理。

为了构造处理基于知识的系统的通用模型，展开对偏差型语义Web服务的自动组合研究。本文针对实际领域所具有的知识多少，在对传统的分析型系统的实例进行研究的基础上又扩展了研究的范畴，将所研究的问题归纳后，定义为偏差问题。

2.1 背景知识

2.1.1 启发式分类

所谓启发式分类，就是通过数据抽象、启发式匹配和细化（refinement）三种操作来对数据进行分类，其中启发式匹配是指将数据启发式地映射到事先列举的解决方案层次上。启发式分类方法包括了两个数据空间领域，提供两个空间的转换映射。与简单分类的不同之处是在不同分类层次上的相关概念是通过非层次、不确定推断得到的。启发式分类是一个计算方法，而不是一个要解决的问题。启发式分类的推理结构如图2-1所示，其中的推理方向、抽象与细化的关系均是简化了的。

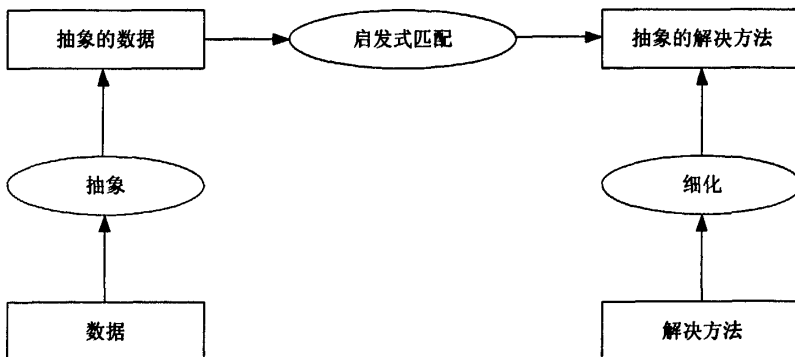


图2-1 启发式分类的推理结构

Figure 2-1 Inference structure of heuristic classification

启发式分类模型主要通过抽象、匹配和细化（refinement）三种操作来解决

问题。

(1) 抽象

数据抽象就是根据领域知识从已有观测值中推理出新的观测值，将一组观测值作为输入、并产生一个观测值作为输出。在数据空间中隐含着一种层次结构：直接观测到的数据是这一层次结构的最底层，高层的数据是通过低层的数据进行抽象得到的。在启发式程序中抽象数据有以下 3 个基本方法：

- 定义性抽象：基于概念的本质且必要的特征。例如：如果结构是一维网络，那么其形状为柱形。
- 定性的抽象：一种含有定量数据的定义形式，通常与一些标准或期望值有关。例如：如果病人是成人且白血球数小于 2500，那么白血球数低。
- 泛化：在子类型层次中泛化。例如：如果客户是法官，那么他是一个受过教育的人。

(2) 细化

解决方案的细化就是以—个解决方案作为输入并产生多个更加具体化的解决方案，即：含有一个现象子集的解决方案可以由父辈的解决方案所解释。在这个过程中要应用到关于解决方案的领域知识。细化引入了一个解决方案空间上的层次结构来表示细化空间。因此细化可以被看作是抽象的逆过程：不是从具体数据到抽象数据，而是从抽象的解决方案到更具体的解决方案。

(3) 启发式匹配

在简单的分类中，数据可以直接匹配解决方案特征或在抽象后匹配。在启发式分类中，解决方案和解决方案特征也要通过另一个分类层次中的一些概念的直接、非层次关联来启发式地匹配。例如，MYCIN 根据生物体的显著特征不仅仅能识别一个未知生物体，而是将生物特征与另一个空间领域建立了联系；MYCIN 启发式地将病人的抽象描述与疾病类别进行匹配。

总之，在启发式分类中抽象的数据描述与特定的问题解决方案或描述解决方案的特征是相关联的。启发式分类揭示了一种在推理模式中存在的普遍行为，是PSMs的思想根源。

2.1.2 问题解决方法

PSMs在当今被看作是构建知识库系统(KBS)的主要组件。PSMs描述了在领域独立的方式下实现一个知识库系统的推理过程，其思想最早源于启发式分类。一个PSMs的描述如下：

- (1) PSMs 明确了为了实现任务的目标需要执行什么样的推理行为。
- (2) PSMs 定义了在这些行为上的一个或多个控制结构。
- (3) 知识角色明确了领域知识在每个推理行为中所扮演的角色。这些知识角色定义了独立于领域的通用术语。主要有两类角色：静态角色，描述了 PSMs 所需的领域知识；动态角色，形成了推理行为的输入和输出。

PSMs 有以下几个目的：

- (1) 知识库系统构建（知识工程）：一个 PSMs 对于构建用于描述实现特殊任务目标的问题解决程序是非常有用的。通常，这里含有一个任务分解方法。
- (2) 知识库系统规范（推理）：一个 PSMs 可以描述一个用于实现任务目标的有效推理过程。在这个意义上，一个 PSMs 关心的是构建过程的产物，这个方法与知识库系统的设计模型相关。
- (3) 认知建模：一个 PSMs 可以描述人类解决问题的认知模型。

在知识工程和知识获取中 PSMs 扮演着重要角色。它可以通过应用领域知识来有效地实现目标，引导领域知识的获取过程，并通过重复使用来推动知识库系统的开发。

2.2 分析型知识系统与偏差问题

基于知识的系统可分为两部分：分析型和综合型，其中，分析型系统一般包括分类、诊断、评估、监测等类型。综合型系统一般包括设计、规划、建模、调度分配等类型。分析型系统通常为针对问题存在的原因查找，即对与标准比较存在偏差的原因进行分析；而综合型系统是在问题分析后对问题进行规划或给出解决方案，但这种传统的分类方法并没有一个统一的处理模型。通过对分析型知识系统的研究，我们发现它们存在一个共同的特征：具有现实值与标准值的偏差，而问题的分析就是发现偏差值中所隐含的信息。因此对分析型系统扩展领域进行研究与归纳后，定义为偏差问题。本文的工作是对分析型系统的研究，而关于综合型问题的研究由另一位同学完成。

2.3 偏差问题

在实际生活中，人们得到的数据值往往与标准值之间存在偏差，我们关注的这类问题的特点是：

- (1) 在具体领域中，存在着一个标准，这个标准可能是自然形成的，例如

医学诊断，健康人就是标准值（隐含）；也可能是一个统计值的集合作为标准，例如河流的水污染监测，一定时间的统计数据就作为标准；也可能是用人们根据经验做出的预测当作标准，例如试卷分析，预期的成绩则可成为标准。这里所述的偏差问题实际包括了基于知识的分析型系统。

(2) 在实际过程中，人们得到的数据往往与标准存在偏差，问题的解决就是要分析出人能够理解的产生偏差的原因。

(3) 所涉及的领域较多，人们对这些领域的知识或多或少，存在不同的解决方法。

(4) 所涉及的数据是结构化数据，数据可是符号型，也可是数字型，数字型可通过离散化转成符号型。

针对实际领域所具有的知识多少和数据偏差的特点，我们将偏差问题分成三种类型：第一类是问题的领域知识比较充分，产生的偏差来自特定的科学规律，一种基本的偏差对应一种情况，如水污染监测；第二类是通常意义下的基于知识的系统，这类问题的领域知识实际上是不够充分的，并没有可依据的科学规律，主要是依靠人类的经验。比如中医诊断系统，知识的来源主要是专家的长期经验的积累。另外，这类问题也可通过机器学习使用由基础数据统计得出的规律。于是这一类问题的知识可采用两种方法加以解决，但后一种方法的处理与第一类基本相同。一般来讲，对于计算机处理，从掌握的信息量来分析，使用学习方法得到的规律要好于使用专家经验，更容易得到人们的信任；第三类是问题的领域知识较少，并不具有类似于第二类那样的与一些偏差对应的层次类别知识。即没有长期经验的积累，也没有大量的基础数据可用来学习规律。只具有一些偏差及其产生原因之间简单的对应关系。例如试卷分析、超市经营分析、球赛数据分析等。

本文的偏差问题是指：存在一类问题，当得到所关心问题的一组特定的已知事实（原始观测值），经过一定的处理在与某一认可的标准比较后，通常将得到偏差值，问题的解决就是要分析出人能够理解的产生偏差的原因。

针对以上偏差问题中的三种类型，根据其特点和相互关系，分别使用适合各自的 PSMs：第一类，我们将其看成是分类模型化方法的特例；第二类，将通过对分类模型化方法的改进和回归模型加以解决；第三类的解决方法，仍使用前期研究的成果，但对细节做进一步的研究和改进。

2.4 偏差问题的问题解决方法

偏差问题的PSMs通过两个空间加以描述，一个是数据空间，一个是解释空间。系统得到的数据存在于一个数据空间，而产生偏差的解释则在一个解释空

间,寻找解释就是在这两个空间中建立联系。PSMs正符合由William J. Clancey提出的启发式分类模型的,因此可以纳入此领域范畴,但是在匹配的方式、抽象与细化的方法上,仍有很大的不同。

数据空间由一个数据集合来描述,用来表示具体问题的事实与标准,以及对初始事实进行抽象或计算后得到的新的数据,但抽象数据目前仍是不具有解释空间特征的。解释空间是由一个解释集合组成,一个解释是空间的一个点,一个解释由能够组成解释的一组解释属性构成。

2.4.1 问题解决方法的三个部分

偏差问题的PSMs的处理过程与启发式分类的思想类似,可分为三个部分:偏差数据、关系对应、解释偏差。

偏差数据对应于启发式分类的抽象部分,但有两点不同。一点是由于引入偏差的概念,此部分又分为三个部分:观测数据、标准数据、偏差计算。对于某些分类或诊断问题,偏差是隐含的,三个部分又合为一个部分,观测数据部分所得结果就是偏差计算部分的结果。另一点是观测数据和标准数据中,对于某些实际领域是通过推理完成数据抽象的工作,而对于另外一些领域则可能是通过公式的计算来完成。

关系对应对应于启发式分类的启发式匹配部分,不同点在于关系对应不仅包括抽象后的启发式关系,而且包括数据的属性与解释的属性之间的启发式关系。也就是说,对应后得到的即可能是一个抽象的解释,也可能是由解释属性及其值组成的集合。对于某些实际问题,偏差数据部分与关系对应部分可合并成一个过程。

解释偏差对应于启发式分类的细化部分。通过解释空间来描述,针对不同的问题,会有不同构造的空间表示,可能是树结构,也可能是表结构,或是多个结构的组合。如果初始数据是具有解释空间属性的,则此问题处理过程只包括解释偏差部分,匹配过程是直接的。

2.4.2 数据空间与抽象

数据空间是偏差问题中数据的集合。在数据空间中,通常涉及五个基本子概念:原始观测值、观测值、原始标准值、标准值和偏差值。其中,观测值和标准值是分别由原始观测值和原始标准值通过推理或计算而得到的。

(1) 观测值

主要的输入数据,包括基础数据值和统计数据值。基础数据值是实际直接观

测到的，即指对一些事物或现象测量所得的原始数据；统计数据值是根据基础数据进行整理得到的。

(2) 标准值

可以是长期观测的平均值，也可以是根据这些平均值在实际需求下加以修改的预测值，还可以是在无历史纪录时通过对一些属性进行分析给出的预测估计值。第一种情况是可以直接获得的，后两种情况是间接获得的。标准值是一种参照量。

(3) 偏差值

将观测值与标准值进行比较，如果存在偏差，则得到偏差值，并需要对它进行解释。因为偏差中通常包括很多潜在而有用的知识，对它的解释可以指导我们的工作。

有时，对于某些简单领域，原始观测值与观测值、原始标准值与标准值都可能是相同的。原始观测值和原始标准值，每一个都可单独构成一个子空间，而观测值、标准值和偏差值可共同构成一个子空间。这些概念都可以被描述为 $\langle f, v \rangle$ 的形式，其中 f 是一个未知对象的属性， v 是这一属性的值。本文称为一个属性值对，并采用一般的观点来看待属性这个概念：属性可以描述一个对象，其值是可以直接观察到或可以推出来的。一组属性值对可以表示一个对象的某一时刻的状态，如果用 N 个属性描述一个对象，则这个空间就是 N 维的，一个对象就是 N 维空间中的一个点。原始观测值是主要的输入数据，一般是实际观测的值。标准值可以是长期观测的平均值，也可以是根据这些平均值经过实际需求加以修改的预测值。偏差值是将观测值与标准值进行比较，若得到的偏差数值在一个允许的范围内，偏差值为空，即没有偏差；否则，偏差值可取偏差数值、大小、高低等值表示偏差。一般出现偏差的属性越少，实际情况与标准或预测越接近，如无一偏差，则完全吻合，无需解释。如果标准值是隐含的，观测值就是偏差值。

启发式分类模型的抽象操作一般可通过规则形式表达，主要有三个基本方法：定义性抽象、定性抽象和泛化抽象，主要通过推理完成操作。偏差问题的抽象概念除了包括启发式分类模型的抽象操作以外，从观测值和标准值的比较中得到偏差值，也看作是一次抽象，如图2-2所示。对于一些领域，观测值与标准值不能直接得到。例如，对于试卷成绩分析，产生的成绩一般呈正态分布，可通过对成绩进行统计得到分布状况，此时的分布特性是所需的观测值，这种抽象可称之为利用已有规律的数学公式的抽象。对于标准值来说，在试卷成绩分析中是通过预测成绩的平均分来体现的。如何预测，原先并没有一个固定标准，分析系统通过综合考虑作为原始标准值的试题本身难度、教学难度和教学要求来给出成绩平均分的预测值。此时的标准值的获得可看成是一种统计的抽

象。

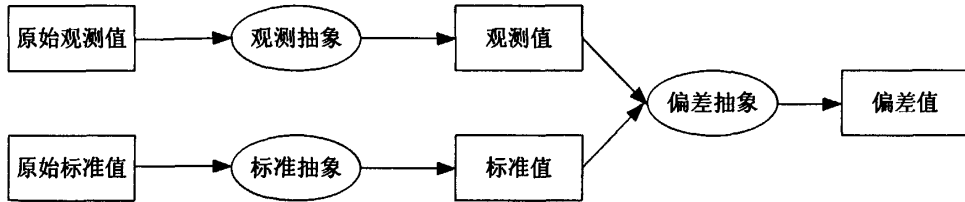


图2-2 数据空间中的抽象

Figure 2-2 Abstract of data space

2.4.3 解释空间与匹配

与数据空间的一组偏差值集合对应的是一个由一组解释值构成的解释集合。因此，解释空间是偏差问题中解释的集合，这个空间的属性值对也可描述成 $\langle g, v' \rangle$ 的形式，并称之为解释值，其中 g 是解释的一个属性， v' 是这个属性的值，一组解释值构成解释空间中的一个点。但由于具体问题不同，偏差问题的解释有时对应解释空间的一个点，而有时是由解释空间的点的有限集合构成。同样由于具体问题不同，所有的解释组成的结构也不同，复杂一点的可以是具有层次的树结构，简单一点的也可以是顺序排列的线性结构。

为了评价根据两个空间属性之间联系而得到的解释能力，需要在偏差值和解释之间进行匹配。实际上，一种匹配模式就是一类问题寻找解释的处理方法。对于偏差问题，将包括下列的多种匹配模式：

(1) 针对偏差问题的第一类问题，由于领域知识充分，匹配关系简单，解释的结构可由线性结构实现，匹配模式是顺序执行的简单模式。对于第二类中可通过机器学习使用由基础数据统计得出的规律的问题，在匹配时，与第一类问题类似，但由于统计的特点，有时会带有不确定的信息，例如：在使用由大量病例统计得到的中医诊断规律时，由于是使用logistic模型计算的，因此带有概率信息。

(2) Enrico Motta和Wenjin Lu提出的分类的模型化方法中的匹配模式是对观测值与解释在同一空间时进行匹配的描述，于是，一个解释可表示为一组属性规范 $\langle g, c \rangle$ 的形式，其中， c 是一个条件或范围。如果 v' 满足条件 c ，则一个解释值 $\langle g, v' \rangle$ 匹配一个属性规范 $\langle g, c \rangle$ 。此模式适合于偏差问题中的第二类问题，即传统的基于知识的分析型任务，解释往往具有层次结构的领域知识的支持，这种模式也适合于启发式分类的细化部分。

(3) 在对上述的偏差问题的第三类问题进行研究的过程中，发现这类问题一般并不很复杂，但对解释没有提供具有层次结构的领域知识，如：试卷成绩

分析、超市经营数据分析、球赛统计数据分析和等领域。这些领域的偏差值与解释值可采用符号型,在它们之间存在着一些简单的对应关系。于是,一组偏差值对应的一个解释值是解释空间的一个点,在没有分类知识的支持下无法分类,即无法在更抽象的层面上说明产生偏差的原因。因此,针对这种情况,在分析了偏差问题的一些实际领域后,总结了两个特点:(1)当偏差值在正常范围内时,则认为无偏差,即这一属性值对被删除,因此,描述偏差的一组属性值对的数目是不定的,不同情况变化可能很大,所对应的解释的属性值对也随之变化。(2)一般情况下,在给定具体的标准后,产生的偏差有两种情况:可定义为正偏差和负偏差。正偏差一般是指对所求问题产生好的或正向的影响的偏差,负偏差则反之。对应到解释上,正偏差产生正解释,负偏差产生负解释。采用一些正解释的解释值构成的表称之为属性值表。根据这一特点,本文采用属性和值分开的方法,通过由解释属性构造的分类层次结构和由属性值对构造的属性值表进行分类类别的构造。由于偏差问题的目的是寻找产生偏差的原因,因此利用属性的分类,得到抽象层面的原因;再通过属性值表的匹配,区分原因所起的作用。偏差问题是通过数据空间产生的不同属性的偏差组合,去寻找不同的解释原因。这样的分类类别,可在更抽象的层面上解释产生偏差的原因。

2.4.3.1 解释属性的分类层次结构

针对第三类问题的匹配模式,在我们的研究实例中,解释属性可分为三种基本类:

(1) 外在的解释属性(即外在原因)

例如,超市经营数据分析领域所涉及到的周边人民生活水平、周边人口等属性。

(2) 自身的解释属性(即自身原因)

可细分为两类,一种是与精神有关、与战略有关或与抽象层面有关的称为隐性的解释属性;另一种是与物质有关、与战术有关或与具体层面有关的称为显性的解释属性。

①隐性的解释属性:如球赛统计数据分析和领域涉及到的运动员心态。

②显性的解释属性:如超市经营数据分析领域涉及到的商品进价和商品售价等。

由此可以看出,由三种基本类的解释属性所构成的解释属性集合共分为8类。比如将外在的解释属性记为a,隐性的自身解释属性记为b,显性的自身解释属性记为c,那么解释属性的8种组合情况如表2-1所示:

表2-1 解释的分类

Table 2-1 Classification of explanation

| 1 类别 | 2 解释属性的组合 |
|--------|-------------------|
| 3 第1类 | 4 $\{\emptyset\}$ |
| 5 第2类 | 6 {a} |
| 7 第3类 | 8 {b} |
| 9 第4类 | 10 {c} |
| 11 第5类 | 12 {ab} |
| 13 第6类 | 14 {ac} |
| 15 第7类 | 16 {bc} |
| 17 第8类 | 18 {abc} |

其中 $\{\emptyset\}$ 代表数据空间中各描述属性值均无偏差的情况（即偏差值均为空），此类无需解释，在分类层次结构中不出现。只有存在偏差的其余7种解释属性组合构成了解释属性的分类层次结构：（如图2-3）

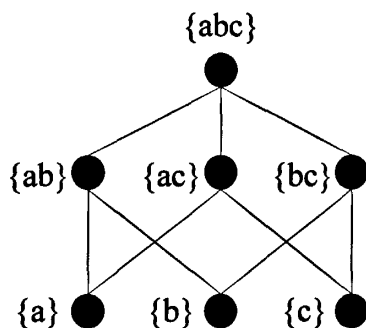


图2-3 分类层次结构

Figure 2-3 Hierarchy structure for classification

2.4.3.2 正解释的属性值表

根据偏差问题的特点，在给定具体的标准后，产生的偏差有两种情况：正偏差和负偏差：

①正偏差：一般是指对所求问题产生好的或正面的影响的偏差；

②负偏差：一般是指对所求问题产生坏的或负面的影响的偏差。

对应到解释上，正偏差产生正解释，负偏差产生负解释。采用一些正解释的解释值构成的表称之为属性值表。

2.4.3.3 解释空间上的匹配

课题组对Enrico Motta和Wenjin Lu提出的分类的模型化方法做出较大修改,给出了新的匹配模式、匹配标准。解释空间上的匹配包括两个方面:属性的分类层次结构上的搜索,属性值表上的匹配。

设一组偏差值 $dev: ((f_1, v_1) (f_2, v_2) \dots (f_m, v_m))$,与其对应的一组解释值为: $((g_1, v_1') (g_2, v_2') \dots (g_n, v_n'))$,一个解释为对属性有要求的一组解释值的集合,属性值表是由一组特定的解释值构成。

在分类层次结构上的搜索:

$(f_i, v_i) \in dev$, 与其对应的是 (g_j, v_j') , 若 $(g_j, v_j') \notin exp$, 则表示 g_j 是不可解释的。

$(f_i, v_i) \notin dev$, 与其对应的是 (g_j, v_j') , 若 $(g_j, v_j') \in exp$, 则表示 g_j 是多余的。

在属性值表上的匹配:

$(f_i, v_i) \in dev$, 若与其对应的是 (g_j, v_j') , 若 $(g_j, v_j') \in$ 属性值表, 则表示 g_j 是可匹配的。

$(f_i, v_i) \in dev$, 若与其对应的是 $(g_j, v_j') \notin$ 属性值表, 而是 (g_j, v_j'') \in 属性值表, 则 g_j 是不匹配的 (v_j' 和 v_j'' 都是 g_j 的值)。

在利用匹配模式进行匹配时,为了得到抽象层次的原因解释及影响程度,还需要确定一个匹配标准,并用它引导在属性的分类层次结构上和属性值表上的匹配。本文用 (I, E, U, M) 的形式表示匹配标准,其中 I 表示不匹配的属性集合, E 是可匹配的属性集合, U 是不可解释的属性集合, M 是多余的属性集合。

2.4.3.4 分类层次结构的搜索

将经过关系对应处理的解释,在解释空间中的分类层次结构上进行搜索。分类层次结构中共包括8类。在分类层次结构搜索后,将得到这个解释所属的类别。分类层次结构的每个节点分别代表一个类别,每个类别包括一些解释属性。在搜索过程中,主要以 U 和 M 为标准来进行比较。匹配标准为:假定两个节点 $S_1=(U_1, M_1)$ 和 $S_2=(U_2, M_2)$, S_1 好于 S_2 , 当且仅当:

$$(U_1 < U_2) \vee (U_1 = U_2 \wedge M_1 < M_2)$$

其中 $X_i < X_j$ 的记法表明 X_i 集合所包含的元素比 X_j 集合少。在分类层次结构上匹配,输入的是与一组偏差值对应的一组解释值,利用匹配模式在以解释属性构建的分类层次结构上进行节点的匹配,并用可比较节点优劣的匹配标准中的 U 和 M ,进行搜索,找到可输出的解释的类别。

搜索是在解释空间中的分类层次结构自顶向下进行的：

- 1) 若顶层的节点好于下一层的所有子节点，则顶层节点就是解；
- 2) 若顶层节点不比下一层子节点好，则取最好的下一层子节点，若最好的有多个，则任取一个。若此时是底层节点，则为解；若不是底层节点，则继续递归搜索。

搜索算法的形式化表示如下：

输入：一个解释所持有的解释属性

输出：这个解释的类别

初始：S 表示一个节点。

将 S 初始化为分类层次结构中的顶层节点。

搜索：Search(S)

```
{
    if(S 是底层节点)
        S 是最优解，得到解释的类别，返回；
    if(S 好于所有直接子节点)
        S 是最优解，得到解释的类别，返回；
    Search(S 的最好直接子节点)；
}
```

2.4.3.5 属性值表上的匹配

在分类层次结构搜索得到解释所属的类别后，下一步就要在解释空间中的属性值表上进行匹配。不同于在分类层次结构搜索中以U和M为标准来比较，在属性值表匹配时，则用到I和E的比较。其中I表示不匹配的属性集合，E是可解释的属性集合。

由一组偏差值得到一组解释值，并将该组解释值按基本类进行区分，然后用每个基本类中的解释属性及其正解释值来构成这个基本类的属性值表。最后将这个基本类中的解释属性的实际解释值与相应的属性值表进行匹配，匹配结果将得到五种可能的情况，分别代表这个基本类对所求问题的不同影响程度：

- ① 所有属性都是可解释的 ($I = \emptyset$)，称之为完全正解释。
- ② 所有属性都是不匹配的 ($E = \emptyset$)，称之为完全负解释。
- ③ 可解释的和不可解释的属性数相等 ($|E| = |I|$)，称之为平解释。
- ④ 可解释的属性数大于不可解释的属性数 ($|E| > |I|$)，称之为基本正解释。
- ⑤ 可解释的属性数小于不可解释的属性数 ($|E| < |I|$)，称之为基本负解释。

在属性的分类层次结构上和属性值表上完成匹配后，将得到解释的类别

和影响程度，以及具体的一组解释值。

2.4.4 关系对应与多对多关系

在数据空间和解释空间之间的联系是通过启发式建立起来的，可以用规则来表达，本文称之为关系对应。这种启发式关系往往由经验而来，建立过程中会忽略掉一些很难被察觉或很难被理解的中间关系。偏差问题的关系对应既包括类似于启发式匹配的抽象对应，也包括偏差值与解释值的直接对应。也就是说，数据空间中的一个属性的偏差值产生的原因可能关系到解释空间的一个属性，也可能是多个属性的；反过来，一个解释空间的一个属性的取值可能是数据空间的几个属性取值的原因。通俗地讲，两个空间的属性取值之间的联系可能是多对多的关系。多个偏差值对应多个解释值的情况，可分解为一对多和多对一的情况。实际上，多个偏差值对应一个解释值的情况产生的问题可由一个偏差值对应多个解释值的情况来解决。对于一对多情况，多个解释值应是一个启发式规则的结论部分，它是由解释值构成的析取式。对于析取式，若能转换成合取式，则可归为一对一的情况。对于多对多的处理，可采用三种方法：

- (1) 若只有一对一的情况，直接使用匹配模式和匹配标准，不做修改。
- (2) 若有一对多的情况，可分两种方法：
 - ① 对于多的方面，可通过获取实际解释属性的值进行确定，不一致的删除，一致的可以将析取改成合取。
 - ② 对于多的方面，直接将析取改成合取，一般是扩大了解释。所得解释是近似的结果。

2.4.5 总结

因此，偏差问题的PSMs如图2-4所示，主要步骤包括三步（抽象、对应、匹配）：首先，输入原始数据，在数据空间通过计算得到观测值与标准值之间的偏差值；其次，根据数据空间中的偏差值，通过启发式推断，得到一个由一组解释值构成的解释；然后，使用解释标准和匹配标准在解释空间中由解释属性构造的分类层次结构上搜索这个解释的类别并在属性值表上区分这个解释的影响程度，完成细化过程，最终得到带有抽象类别的人易理解的偏差解释。此方法我们已经针对试卷成绩分析、超市经营数据分析、球赛统计数据领域进行了实例研究^[18]，并取得了好的结果。

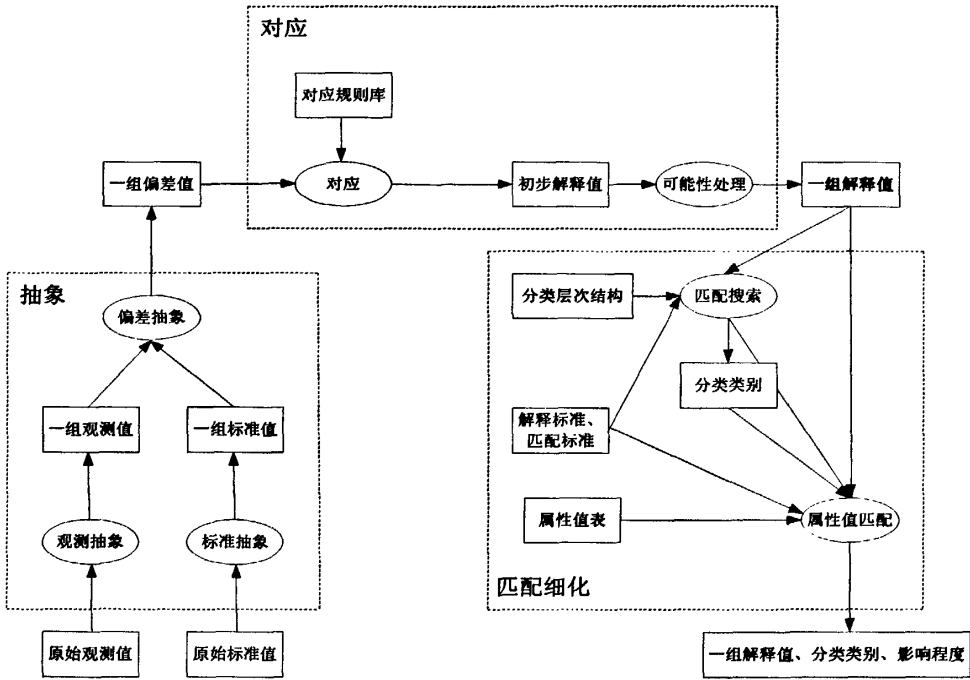


图 2-4 偏差问题的 PSMs

Figure 2-4 Problem solving methods of deviation problem

2.5 本章小结

本章给出了偏差问题定义和其研究范围，及偏差问题的三种分类和各自特点，并提出了对应的PSMs。PSMs主要包括三个部分：抽象，对应和匹配。在偏差问题和PSMs的论述中，穿插了对偏差问题的特征进行讨论，及各不同情况下PSMs的改变。

第3章 偏差型语义 Web 服务

对偏差问题的 PSMs 进行研究,分析各概念之间的关系,研究如何将 PSMs 分成多个小粒度语义 Web 服务的组合,给出了解决偏差问题的语义 Web 服务组合架构,称为偏差型语义 Web 服务。

3.1 语义 Web 服务

World Wide Web (WWW) 是一种在TCP/IP协议之上的特殊结构框架,通过链接机制可以访问遍布全球的Internet资源。有关WWW的各种技术通称为Web技术。当前Web技术的体系结构主要是基于HTML (Hyper-Text MarkupLanguage) 和HTTP协议。随着Web迅速地普及,基于Web的应用也由最初的简单应用延伸到种类日益繁多的复杂应用和计算。近年来,WWW的发展出现了两个新兴的重要发展趋势。

一个是Web服务 (Web Services)。Web服务是一些自包含、自描述、模块化的程序,可以发布在Web上,并被发现和调用。Web服务使得WWW从静态的、松散的Web页面集合逐渐演化成动态的、互联的应用和服务平台。其接口和绑定可以通过XML构件进行定义、描述和发现,Web服务支持通过基于因特网的协议使用基于XML的消息与其他软件应用程序直接交互。Web服务所执行的功能可以是简单的请求,也可以是复杂的商业过程。服务可以提供信息,例如天气预报、股票报价服务;也可以对现实世界产生影响,例如机票预订、信用卡交易等服务。越来越多的服务提供商开始基于Web向用户提供服务,例如B2B应用、B2C应用等。Web服务的体系结构是基于三个角色 (服务提供者、服务请求者、服务中介) 和三个操作 (发布、发现、绑定) 构建的。服务提供者通过在服务中介处注册来发布可用的服务,服务请求者通过在服务中介处查找服务登记记录来发现合适的服务,并绑定服务提供者以使用该服务。Web服务的角色及操作如图3-1所示。

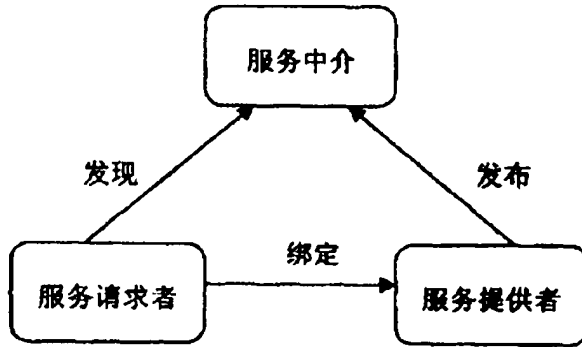


图3-1 Web服务的角色及操作

Figure 3-1 Roles and Operation of Web services

Web服务是由一套协议栈构成的层次化体系结构,如图3-2所示,其中底层传输协议是广泛使用的因特网标准:HTTP, FTP, SMTP等。中间部分是目前Web服务的相关协议标准,包括简单对象访问协议SOAP (Simple Object Access Protocol)、统一描述/发现/集成协议UDDI (Universal Description, Discovery and Integration), Web服务描述语言WSDL (Web Service Description Language)。SOAP是一种基于XML的不依赖于传输协议的表示层协议,用于应用程序以对象形式方便地交换数据。UDDI是一套信息注册中心的规范,所有Web服务注册信息都存储在UDDI商业注册中心,以便查找。WSDL通过一套XML语法描述,将Web服务定义为一个能交换消息的通信端点集合。在中间层之上,是一些支持复杂工作流和业务逻辑的工作流建模语言。图3-2右边部分是各个协议层的公用机制,包括安全和管理的协议或机制。

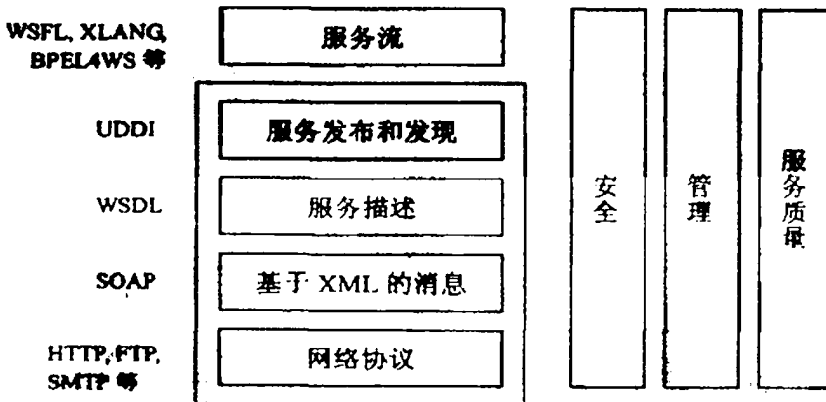


图3-2 Web服务的体系结构

Figure 3-2 Structure of Web services

然而,目前Web服务的描述都是基于半形式化的自然语言,如UDDI, WSDL等,只支持手工检索和手工组合,对服务的自动发现、协商、组合等的支持还很有限。

另一个是语义Web (Semantic Web)。语义Web的概念由Tim Berners-Lee于1998年首次提出。语义Web研究的目的是扩展当前的Web赋予Web中的所有信息以定义良好的语义,让计算机能够理解和处理,从而使人和计算机能更好地协作,充分发挥WWW这个巨大的分布式信息系统的潜力。

语义Web研究的重点是如何把信息表示为计算机能够理解和处理的形式,即带有语义的形式。Berners-Lee在XML2000国际会议上提出了语义Web的层次结构,如图3-3所示。它主要基于XML和RDF (Resource Description Framework) / RDFS (RDF Schema),并在此之上提供本体和逻辑推理规则,以完成基于语义的知识表示(KR)和推理,从而能为计算机所理解和处理。语义Web在理论上以知识工程(KE)成熟的理论、技术和方法为核心,知识表示、ontology和智能agent等都是重要的研究内容,它们是不可分割的。

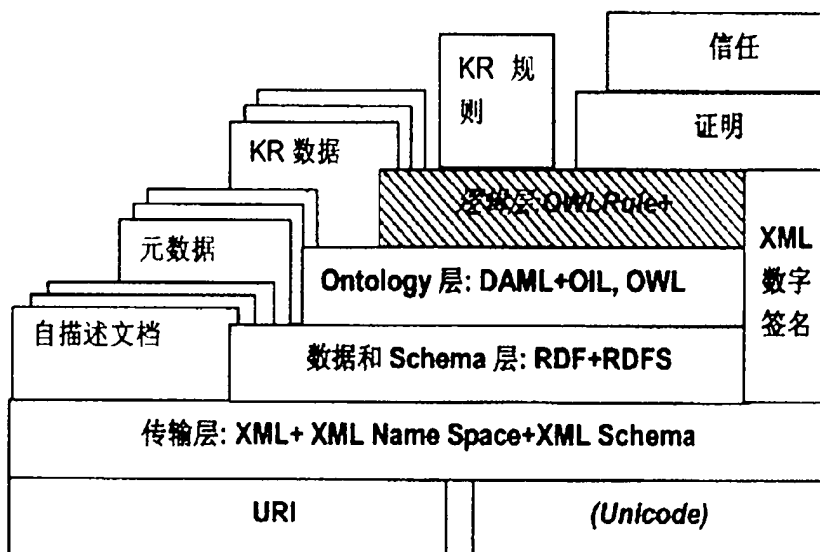


图3-3 语义Web的层次结构

Figure 3-3 Layer structure of semantic web services

3.2 偏差型语义 Web 服务

本章将偏差问题的PSMs转化为语义Web服务的组合架构,并称之为偏差型语义Web服务。首先提炼出PSMs的基本概念,通过对概念的分析将找到PSMs对应的组合架构,归结一定粒度的服务,然后再将其转化成Web服务组合。下面将通过偏差数据、关系对应、解释偏差、观测数据、标准数据、偏差计算、分类层次结构搜索、属性值表匹配等概念以及这些概念之间的关系加以说明。

3.2.1 偏差问题的语义 Web 服务组合架构

偏差问题的组合架构是在PSMs的基础上提出来的，描述了偏差问题的处理过程，如图3-4所示。

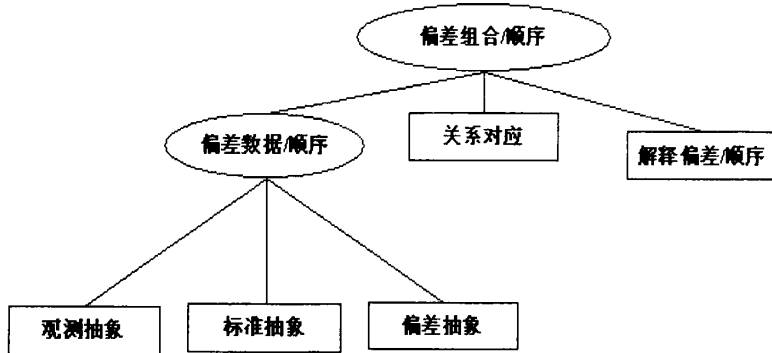


图3-4 偏差问题的语义Web服务组合架构

Figure 3-4 Composition framewrok of semantic web services of deviation problem

对于一个具体的偏差问题在输入了原始观测值和原始标准值后，通过组合架构执行将得到关于产生偏差的解释。具体的架构分为三个部分，共七个过程。

(1) 第一部分为抽象部分，包括观测抽象、标准抽象和偏差抽象过程。

①观测抽象：由原始观测值得到所需观测值的过程。

②标准抽象：由原始标准值得到所需标准值的过程。

③偏差抽象：通过对观测值和标准值进行比较得到偏差值的过程。

(2) 第二部分为对应部分，包括简单对应和可能性处理过程。

①简单对应：输入数据空间中的一组偏差值集合，根据对应规则库得到一组未定的解释值。

②可能性处理：在进行多对多关系的处理后得到一组具体的解释值。

(3) 第三部分为匹配部分，由于解释的结构不同，使用的匹配方法也不同，对于具有层次的结构，可使用结构匹配，对于线性结构，可使用简单匹配。例如偏差问题的第三种，使用两种匹配的组合，即对于分类层次结构的一种结构匹配和对属性值表的一种简单匹配的组合。

针对一个实际问题可能只包括部分过程，有时第一和第二部分的界限并不清晰，有时问题只包括第三部分。所以，实际问题的语义Web服务组合结构是多种多样的。例如水污染系统中通过观测抽象和标准抽象得到偏差数据后，使用一次规则即可得到解释结果，也就不包括匹配搜索部分和属性值匹配部分。中医系统中使用丰富的专家经验进行判断，即将得到的偏差数据进行关系对应后即得到解释结果。因此水污染的组合架构包括偏差数据和关系对应部分，而中医诊断的组合架构包括三个部分，但解释偏差部分中不再包括分类层次结构

搜索和属性值表匹配。

3.2.2 组合架构的分层树状结构

偏差型语义 Web 服务由一系列语义 Web 服务的逐层组合来实现其基于知识系统的功能。通过对一批基于知识的系统的分析，以系统具有的领域知识为标准归纳出各自的用本体描述的 PSMs，并将这些方法细化、分层拆分、转化成语义 Web 服务的组合。

一个待解决的实际偏差问题可称之为一个偏差任务，可由一个偏差型语义 Web 服务来描述。偏差型语义 Web 服务或由一个具体的服务或低一层的服务组合而成，自顶向下，逐层分解。因此，偏差型语义 Web 服务的组合架构是呈树状结构的，本文称为语义 Web 服务组合树。我们定义树中的节点分为两种类型：知识级语义 Web 服务、原子级语义 Web 服务。知识级语义 Web 服务用于将任务分解成多个子任务，一个知识级语义 Web 服务可完成一种在知识层面操作的功能，而原子级语义 Web 服务用领域知识的假设进行某操作，实现一定粒度的任务所要求实现的功能，一个原子级语义 Web 服务可完成一种基本操作的功能。不同问题的根节点可通过不同的知识级语义 Web 服务来分解，因此从根中引出的虚线对应不同的解决路径，表示不同的分解方法。当子任务粒度分解到足够小时，就可用某一原子级语义 Web 服务来解决。组合架构树中的语义 Web 服务可以是抽象高层次的，例如诊断或设计任务，也可以是具体的可以解决某一细粒度任务的服务。

因此，在偏差问题的语义 Web 服务组合树中，最高层由图 3-4 的服务组合构成。其中的叶节点可由知识级语义 Web 服务和原子级语义 Web 服务的组合构成，而知识级语义 Web 服务也是由原子级语义 Web 服务组成。

知识级语义 Web 服务是在通常意义下执行知识的推理、匹配以及有关计算的服务，在偏差问题的服务组合中包括经典的正向推理、反向推理、简单匹配、树匹配等类型。

原子级语义 Web 服务是服务组合中的最基本的元素，在 WSMO 体系中实现为一个 LISP 函数，包括初始化、选择参数、扩展节点、取节点、节点匹配等服务，也包括实现控制功能的服务。

3.3 Web 服务组合控制算子

如何在组合架构中管理服务间的关系，包括定义每种执行顺序（顺序、选择、循环等）的定义和其对应的执行步骤定义，这是将组合架构转换为一个可执行

工作流的前提基础，也正是本节讨论的内容。因此，为了在组合架构中加入反映服务组合关系的控制，我们扩展了组合架构而加入了组合控制算子，如图3-5所示。

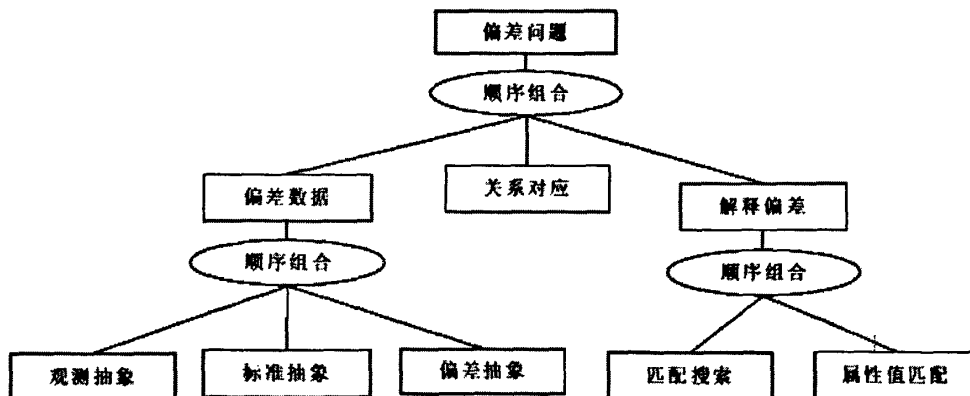


图3-5 扩展的服务组合架构

Figure 3-5 Expanded framework of semantic web composition

扩展后的偏差问题的服务组合架构由执行节点和控制节点组成。图3-5中方形为执行节点，椭圆为顺序控制节点，另外椭圆还可表示循环控制节点。菱形表示条件节点。根节点和中间节点均是控制节点，叶节点为执行节点，表示一个具体的服务或低一层的服务组合。顺序服务控制节点由图中点的位置默认表示，定义为从左到右执行。执行完最右边的一个节点为完成这个组合服务。控制算子结构包括：

- 顺序 (Sequence)：由图中点的位置默认表示，定义为从左到右执行。执行完最右边的一个节点为完成这个组合服务。
- 选择 (If)：由一个条件节点和子节点组成，表示如果满足条件则进行子服务的组合执行，然后继续执行右边的节点。如果不满足，则表示本if节点处理完毕，继续执行右边的点。如：如果库存中找到书，则进行交易和寄书，且交易和寄书前后执行，组合成一个服务。
- 循环：转换为一个循环节点，由一个if子节点和一个Sequence子节点组成。While为if条件在循环的最左边，Do-While为if条件在循环的最右边，表示的意义为：如果满足if条则执行下面的，如果不满足则表示本if节点处理完毕，继续执行右边的点。执行最右边节点默认进行最左边的节点，至到退出节点为结束本层循环，表示父类的循环节点完成，再执行父类的右边节点。

3.4 组合架构的具体应用

下面介绍组合架构在试卷分析^[18]中的具体应用，偏差数据和关系对应部分

与图3-4相同。解释偏差部分的核心是在明确了解释空间的构造后，根据问题的需求采用什么样的匹配算法。试卷分析实例属于前述第三类的偏差问题，根据此类问题没有成熟的层次类别结构，采用属性和属性值分开匹配的方法，先在以属性构建的树状的分层次结构上进行树匹配，再用所得类别结果在解释属性值表上进行简单匹配，做进一步的区分。因此，试卷分析实例的解释偏差部分是一个树匹配和简单匹配的服务组合，如图3-6所示。



图3-6 扩展的试卷分析的解释偏差

Figure 3-6 Explain deviation of expanded paper analysis

下面以试卷分析中的树匹配为实例，使用与图3-5中组合架构同样的执行定义得到图3-7的树匹配服务组合控制图。

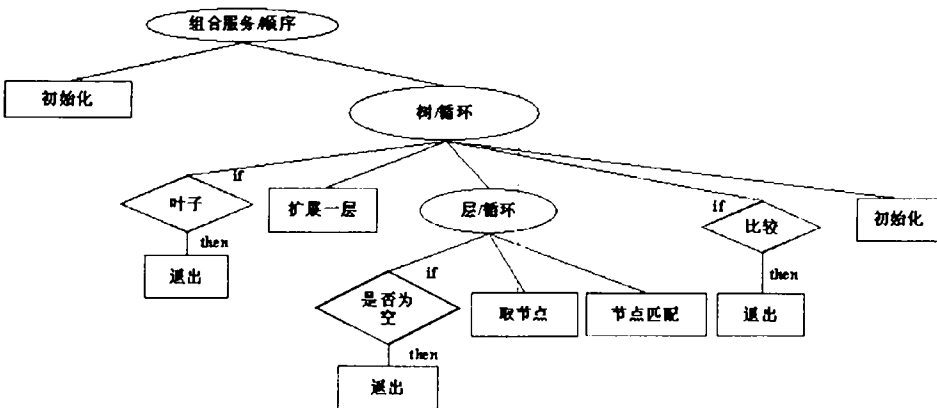


图3-7 树匹配服务组合执行控制

Figure 3-7 Execution control of tree match composition

图3-7中使用的节点类型包括：由一菱形If条件节点和Then执行节点组成的是条件控制节点，表示如果满足条件则执行Then节点，然后继续执行下一步。如果不满足，则表示本条件节点处理完毕，直接执行下一步。椭圆循环控制节点，从结构上可看成是一个顺序控制节点，但必须至少包括一个条件控制节点。当While为条件时，条件控制节点在循环的最左边，Do-While为条件时，条件控制节点在循环的最右边，循环控制节点表示的意义为：顺序执行，执行到最右边节点后默认再执行最左边的节点，直至通过条件控制节点退出为结束本层循环，底层的循环节点完成，再执行上层下一步。

关于图3-7表示的树匹配服务组合的执行过程说明如下：

- 首先进行整个服务的初始化操作，然后进入树匹配查找，在此为一个循环操作，每一个循环操作树中的一层节点。
- 进入树匹配查找，首先判断当前节点是否为叶子节点，如果是则退出循环，否则认为已对if节点处理完毕，且不执行then节点，则默认if节点处理完毕，继续循序执行后面的节点。
- 当前节点不为叶子节点，则扩展当前节点的下一层节点，然后进入下一层节点的循环。
- 进入层循环，首先判断当前层节点表中是否为空，如果是则退出本循环，否则认为已对if节点处理完毕，且不执行then节点，则默认if节点处理完毕，继续循序执行后面的取节点操作。
- 从当前层节点列表中取一个节点。
- 进行节点匹配，计算匹配度。然后循环到层循环节点的第一个节点继续。
- 当层节点表为空时，表示节点处理完毕，则退出层循环。继续顺序执行树循环中右面的节点。
- 对层循环中的匹配度进行比较，判断时候有下层节点优于当前节点，如果没有，则当前节点为结果，则退出树循环。如果存在，则选择此节点为当前节点，且认为已对if节点处理完毕和不执行then节点，则默认if节点处理完毕，继续循序执行后面的节点。
- 进行树循环的初始化操作。然后循环到层循环节点的第一个节点继续。

此组合操作的树循环一共有两个退出口，其一为当前节点为叶子节点，无法继续扩展；其二为找到最优节点，无需继续扩展。

3.5 本章小结

本章基于偏差问题的PSMs，给出了偏差问题的语义Web服务的组合架构。此架构是第四章实现偏差型语义Web服务的指导方法，同时也为第五章中的服务组合的可执行工作流的转换提供了理论和实现基础，而且本章中的控制算子是第四章中偏差型语义Web服务可执行的理论基础。本章中讨论的组合架构是分层树状结构的，而在第五章中将讨论如何把树状结构转换为线性可执行工作流。

第4章 偏差型语义 Web 服务构建与发布

根据第三章的组合架构,本章讨论将偏差型语义Web服务采用本体描述,基于WSMO(知识系统的语义Web服务构建平台),实现偏差问题的实验性可在网上发布的语义Web服务组合架构。

试卷分析系统的PSMs由解释属性构造的分类层次结构和由属性值对构造的属性值表构成,而其对应的服务组合架构中为一个树匹配和简单匹配的服务组合,如图3-6所示。由于试卷分析系统的组合架构是一个比较能完整包含偏差问题PSMs各概念的实例,因此我们选择试卷分析系统作为构建实例。我们已经在WSMO上实现了试卷分析的Web服务组合架构,但由于其复杂性,本文只给出其本体描述和服务结构图。而以分类层次结构匹配中的树匹配的语义Web服务的实现为例,讲述使用WSMO等构建偏差型语义Web服务系统的过程,并加入了服务组合控制算子描述。

4.1 语义 Web 服务的 WSMO 平台

当前语义Web 服务建模框架主要有三种:OWL-S (Web Ontology Language for Services)、IRS和WSMF (Web Service Modeling Framework)。OWL-S 基于面向代理的思想,WSMF 则是面向业务的方法,主要着眼于解决包括信任与安全机制在内的电子商务领域的应用需求,而IRS是从"可重用的知识组件"方法演化而来,属于知识工程方法。因此我们选择使用IRS、WSMO来构建我们的偏差型语义Web服务。

WSMO平台可以很好的支持对本体的管理(概念的相逆、对称、传递、继承和对概念间特定条件的约束等),充分反映概念间的实际存在关系,并使用约束机制在运行过程中对概念间的关系进行动态约束和维护。在WSMO中,顶层的组件包括目标(Goal)、Web Service、本体集和桥(Mediator)。本体介于其它三个组件之间,支持语义的互操作,目标表达请求者的意图,Web Service反映了Web服务的真实行为,桥确定互操作机制和描述性的连接以上的组件,处理工程中产生的不匹配。桥组件有四种类型:OO-桥连接本体与本体;WW-桥连接Web服务与其它Web服务;WG-桥连接目标与Web服务;GG-桥连接目标与目标。WSMO Studio是由WSMO小组开发的一个开源的语义Web服务和语义模型本体构建平台,本文的实例都使用WSMO Studio 0.4.1实现,并通过WSMO内嵌的IRS插件进行发布。

在WSMO中构建的目标,Web Service,桥和本体模型都使用WSML语言描述,通过IRS进行发布后生成相应的OCML描述。具体的Web服务使用Common

Lisp语言编写，Web Service通过WSMO中的Grounding^[31]技术与具体的Web服务（Common Lisp）建立数据通道，并通过IRS-III发布后提供其它服务的语义查找。但是由于目前WSMO提供开源使用的版本仅是实验性版本，因此在有些部分还具有一定的限制：部分功能没有完全公开，如对语义服务组合的支持；部署服务与语义发布描述之间的数据通道的封闭性，再加上WSMO目前并没有一个完整的使用手册，数据传递的调试不很方便等原因，最终只能在Common Lisp环境下模拟实现。

4.1.1 UPML 框架

IRS、WSMO模型起源于UPML框架。UPML所描述的知识库系统包括六个不同的组件：本体（Ontology）、任务（Task）、问题解决方法（PSM）、领域模型（Domain Model）、桥（Bridge）和细化器（Refiner）。

- (1) 任务：定义了知识库系统所要解决的问题。
- (2) 问题解决方法：定义了一个知识库系统的推理过程。
- (3) 领域模型：描述了知识库系统的领域知识。
- (4) 本体：本体提供了在任务、问题解决方法和领域模型中的定义所要应用到的术语。
- (5) 桥是明确的模拟一个体系结构中的不同部分的关系，比如：模拟领域和任务的关系或者模拟任务和问题解决方法的关系。
- (6) 细化器可以被用来表述将组件逐步的具体化而适应其他组件的过程，比如：任务可被细化或问题解决方法可被细化。

UPML的体系结构如图4-1。本体组件被任务、问题解决方法和领域模型三个组件用来定义它们所讨论的领域，图4-1中由表示本体的正方形的粗箭头就表示这种关系。每一个组件都可以通过细化器来具体化到某一个特定的环境。桥表示对任务、问题解决方法和领域模型组件之间的相互关系的控制。在图中用双向的箭头表示这种控制关系，而具体的桥名写在箭头上。

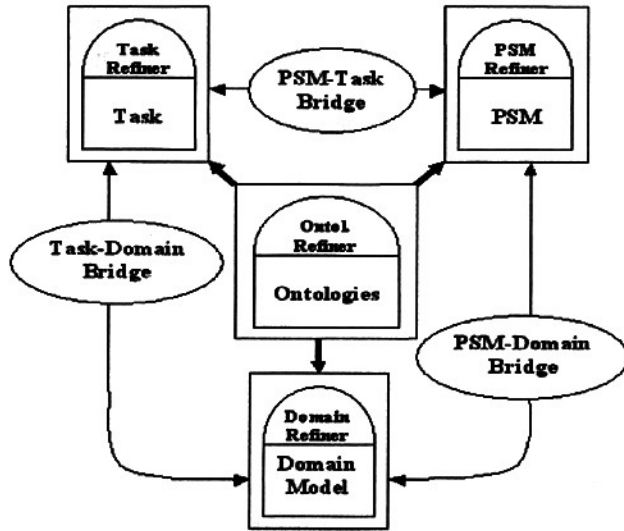


图4-1 UPML的体系结构图

Figure 4-1 Structure of UPML

4.1.2 基于 UPML 框架的 IRS 模型

IRS模型是英国Open University提出的支持Web Services语义理解和语义组合的一个尝试。它将基于知识的方法应用于Web Service的检索和组合之中，并以UPML框架为基础，显示地刻画了领域知识（即领域任务模型）和抽象的问题解决方法，并通过一定的推理手段来实现任务和问题求解方法之间的映射。IRS首先提出基于任务本体的Web Service发布，发现和组合方法，可以说是将Web Service发现和组合从语法的层次提升到了语义的层次。IRS是以UPML为基础构建起来的，其主要包括以下几个主要组件（图4-2）：

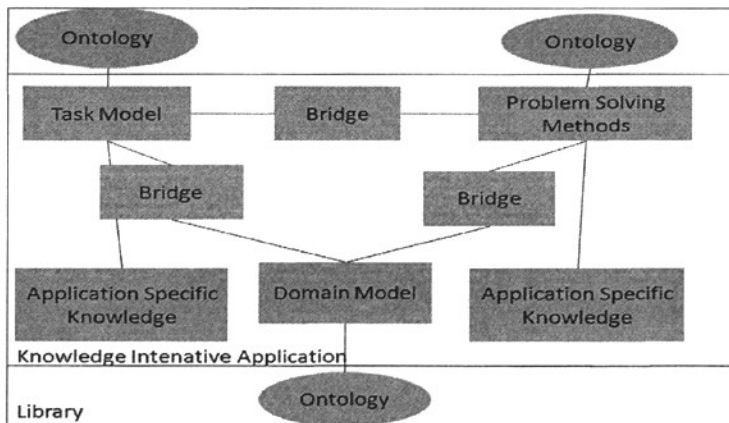


图4-2 IRS的UPML框架

Figure 4-2 UPML framework of IRS

- **领域模型 (Domain Model)**
领域模型描述了定义问题解决方法和任务所需要的领域知识。
- **任务 (Task Model)**
提供对要解决的问题的描述,特别是输入和输出数据类型,任务要达到目标和前提条件。任务可以是一种高层次的对象任务,如分类和模型;也可以是具体的任务,例如汇率兑换任务。
- **问题解决方法 (PSMs)**
它提供抽象的独立的推理过程的描述,使其可以应用于特定领域中的任务。它可以是一个高层抽象的PSMs,如启发式分类。也可以是一个特定方法对应于一个细颗粒任务,如汇率兑换变化。
- **桥 (Bridges)**
它可以建立不同模型之间的映射,商业异类分布式组件或Web服务都是使用桥来解决数据和过程运行中的不匹配。

4.1.3 WSMO (Web Service Modeling Ontology) 体系

在WSMO中,顶层组件包括目标 (Goal), Web 服务 (Web Services), 本体集 (Ontologies) 和桥 (Mediators), 如图4-3所示。

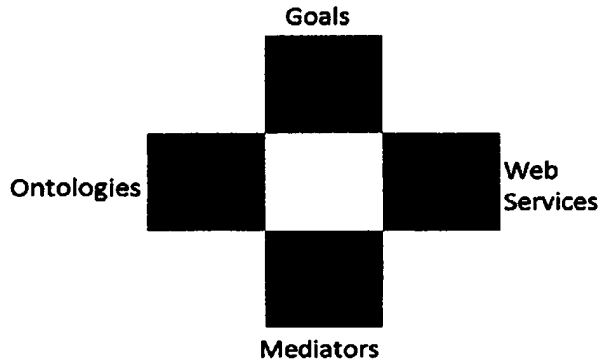


图4-3 WSMO体系结构

Figure 4-3 Architecture of WSMO

- 本体介于其它三个主件之间,支持语义地互操作。它提供了语义描述数据的功能,以达到完成语义互操作和被其它三个WSMO组件使用。
- 目标描述了用户希望完成的功能和执行某给定Web服务的前提条件和事后状态。这样,可以表达请求者的意图。
- Web Service反映了Web服务的真实行为,包括功能和接口,接口包括两个密切相关的概念:Choreograph和Orchestration^[32-35],前者描述与Web服务联系所需的信息,后者为包含合成Web服务的描述信息。后者提供了

执行所有服务组合的所需的所有的细节，并且作为提供服务者的私有项，其它人不可以访问。

- 桥确定了互操作的机制和描述性的连接以上的组件。特别是，WSMO 提供了四种不同类型的桥，支持三种不同的异构处理过程：数据桥，目标桥和过程桥，如图4-4所示：

- ① OO-桥：使组件输出异类本体。
- ② WW-桥：Web服务与其它Web服务连接。
- ③ WG-桥：连接目标与Web服务。
- ④ GG-桥：连接目标与目标。

Choreography描述如何与一个已部署的Web服务相联系。在语义上，Choreography由一系列前向链规则和一个OCML的Grounding声明表示的。当相关条件满足时，一个规则执行通信单元中一些连接行为。Grounding声明涉及通信单元的操作和相关的映射。特别地，每一个输入和输出操作都与一个上升或降低函数相关联。同样地，Grounding也涉及关于相应的发布平台的信息。这种方法允许Web服务的功能通过调用一个或多个声明的操作被识别。

Orchestration用于描述Web服务的组合。在语义上，Orchestration被表示为OCML中的工作流模型。这个模型的特点是它的组合的基本单元是目标。因此，此模型提供了控制和一系列目标的数据流构造，而且，数据流和目标间不匹配的解决由桥来支持。

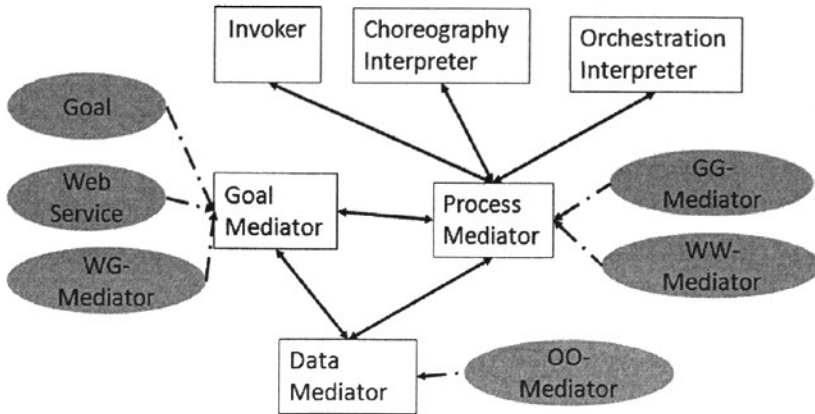


图4-4 桥框架体系

Figure 4-4 Framework of mediator

4.1.4 IRS 发布平台

IRS-III是实现了WSMO描述的一个框架和语义Web服务实现体系。它支持语

义Web服务的发布,调用和组合。IRS-III可自动将程序代码转化为Web服务,并支持功能驱动功能的发现和调用。另外,从其它Web服务的角度来看,任何发布在IRS-III上的服务可以自动地表现为标准的Web服务。IRS-III的主要组件包括IRS服务器、IRS发布器和IRS客户机,它们通过基于SOAP的协议进行通信(如图4-5)。

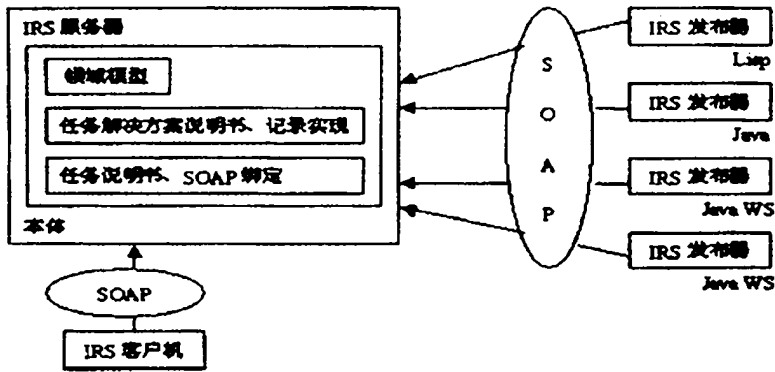


图4-5 IRS-III 体系结构

Figure 4-5 Architecture of IRS-III

IRS服务器主要将功能说明书映射到特定的Web服务。IRS服务器有两个不同层次Web Services语义描述:知识层次的描述和映射机制描述。服务器中的本体构成了第一层次知识层次的描述,使用OCML语言。第二层次的映射机制描述将服务功能描述和具体的Web Service 对应起来,使用DAML-S中的Grounding概念。

IRS发布器主要功能:首先将一段独立的Java或Lisp代码转化为标准的Java Web服务,再将Web服务链接到IRS服务器中相应的语义描述信息;IRS发布器有两项功能:一是关联Web Services和其存在于IRS服务器中的语义描述。一条语义描述可以对应多个实现该描述的服务,一个服务也可以对应多条语义描述。二是为单独的java或lisp代码生成一系列包装,使它们发布成由问题解决方法抽象成的Web Services。发布器可以为这样的代码自动生成Web Services端点。在IRS发布器上发布标准的Web Service,使用如图4-6的界面。

用户只需通过IRS客户机提出一个任务即刻触发相应服务。IRS客户端有一个重要特征:它是功能驱动的。IRS客户端提供一个界面和一系列有关任务的API来支持功能驱动。



图4-6 IRS-III 发布一个服务

Figure 4-6 Publish of web through IRS-III

4.1.5 语义 Web 服务描述语言-WSML

WSML (Web Service Modeling Language) 是 Web 服务模型化语言, 它是由 WSML 工作小组开发的一个基于规则的适合于语义 Web 服务的模型化语言。WSML 基于 WSMO 模型开发, 因此其主要的组件同样包括目标, Web 服务, 本体集和桥。WSML 语言中各组件的描述提供了 WSMO 模型中的各组件定义的实现。目标描述, 如图 4-7 所示:

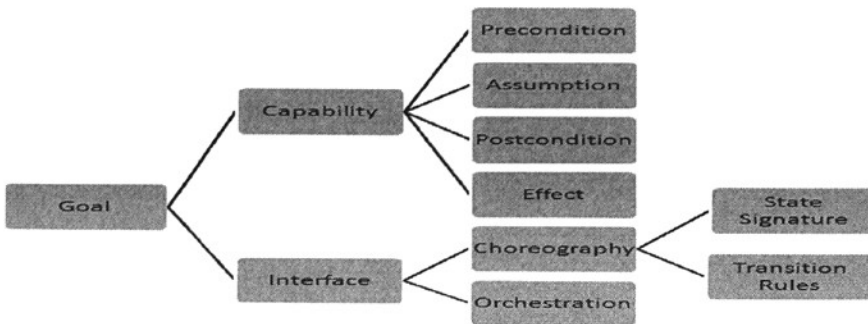


图4-7 目标的WSDL描述

Figure 4-7 Goal description using WSDL language

- **Capability:**描述功能性需求。
- **Precondition:** 用户可以使用服务的保证, 表达了用户与服务交换信息的前提条件。
- **Assumption:** 用户可以使用服务的保证, 包括如何连接和环境。
- **Postcondition:** 用户使用服务后所需满足的条件。

- **Effect:** 用户可以使用服务的保证, 使用服务后用户所希望达到的状态。
 - **Interface:** 描述用户的请求行为, 如连接中的请求约束。
 - **Choreography:** 表达了如何与单个服务建立联系。
 - **State Signature^[36]:** 通过使用modes语义的描述连接的建立。
 - **IN modes:**描述用户期望的连接。
 - **Out modes:**描述服务的返回。
 - **Transition Rules:** 用于约束连接的行为。
 - **Orchestration:** 描述多个Web服务的组合成当前的服务。
- Web服务:** 描述服务信息。其组成部件的含义与目标描述一样, 如图4-8所示。

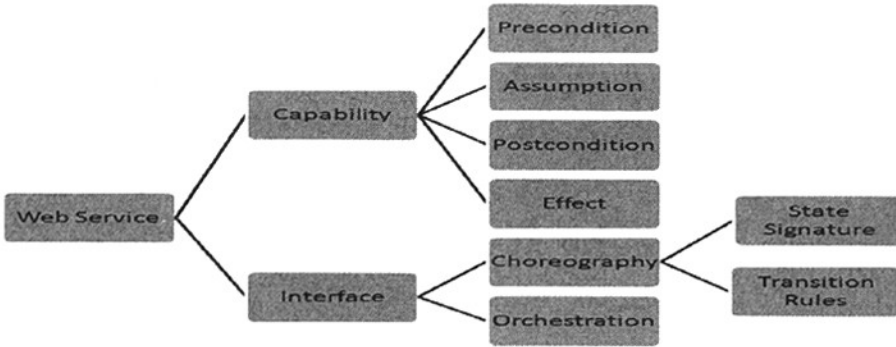


图4-8 Web服务的WSDL描述

Figure 4-8 Web services description using WSDL language

4.2 控制算子的 OCML 描述

WSMO Studio目前未开放服务组合功能, 将服务组合功能扩展嵌入到WSMO与IRS系统可以在两个层面上进行: WSMO的组合扩展, OCML组合扩展。由于OCML是WSMO和IRS的底层语言, 而且WSMO描述的服务的语义通过IRS发布后转换为OCML描述等待其它服务的调用。因此, 对于第三章中的服务组合控制算子的扩展使用OCML语言描述。

扩展的定义包括根节点, 服务描述, 顺序, 选择, While循环, DoWhile循环。服务描述包括服务能够被调用条件的描述(一般为输入要求得到前一个的输出, 并其值满足此要求)和一个绑定具体服务的接口。While循环节点包括三个槽: 选择, 顺序, 退出, 而DoWhile循环包括三个槽: 顺序, 选择, 退出。

根节点: Structured-orchestration

 Has-root-components: value

服务描述: Service-component

 Has-invocable-description: value

```

Has-internal-binds: value
循序执行: Sequence
Has-components: value
选择判断: IfThenElse
Has-if-components: value
Has-then-components: value
While 循环: While
Has-ifthen-components: value
Has-sequence-components: value
Has-return-components: value
DoWhile 循环: While
Has-sequence-components: value
Has-ifthen-components: value
Has-return-components: value
    
```

4.3 偏差问题的语义 Web 服务实例

本节以试卷分析作为实例构建偏差型语义 Web 服务，其本体集如图 4-9 所示。在 WSMO 中，根本体定义为本体，其中 Concept 概念为本体定义（以下都将 Concept 概念直接称为本体）。试卷分析^[18]语义 Web 服务的本体集包括十个本体定义：偏差值，标准值，数值匹配度，节点，节点匹配度，原始值，解释，隐性，显形和外在。其中隐性，显形和外在本体属于节点本体的子本体，而原始值有四个本体实例，分别为四个学生的原始成绩。

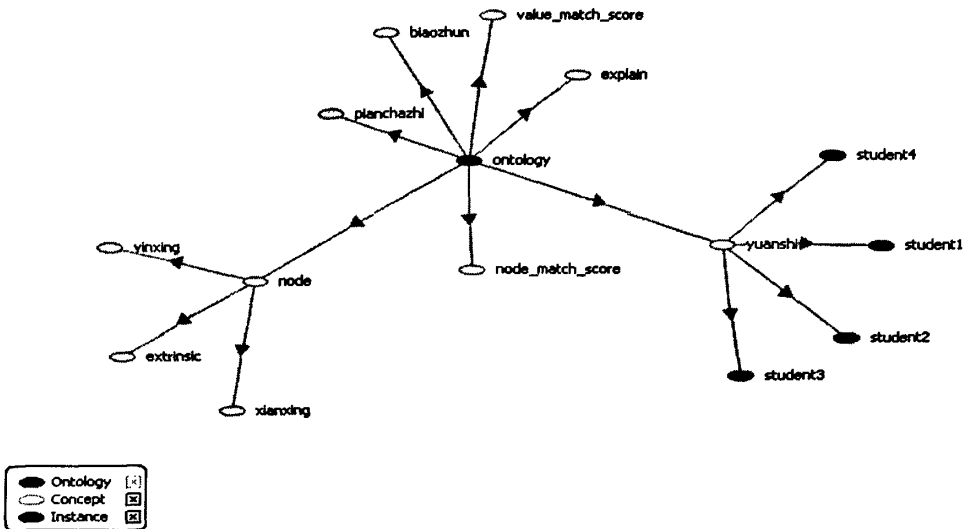


图4-9 试卷分析系统本体集

Figure 4-9 Ontologies of paper analysis

试卷分析的PSMs反映整个服务体系，是我们用于处理知识系统的关键框架，如图4-10所示。

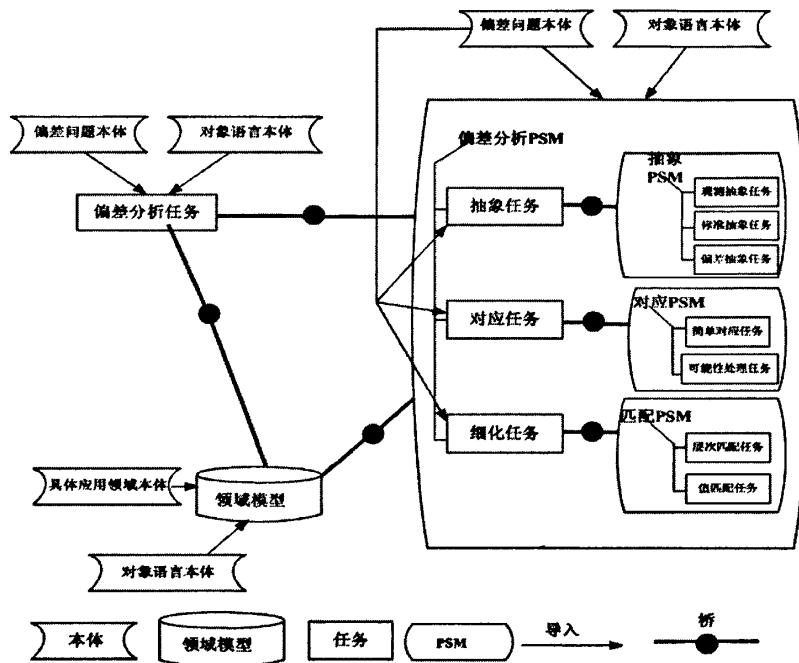


图4-10 试卷分析的PSMs描述

Figure 4-10 PSMs of deviation problem

4.4 分类层次结构匹配的语义 Web 服务实例

偏差型语义Web服务的构建主要包括四个步骤：1)使用WSMO构建本体。2)在WSMO中，使用本体描述，构建相应的服务描述，目标描述及桥描述。在WSMO中将自动转换为WSML语言，也可以直接使用WSML编写。3)通过IRS发布到IRS服务器上，使其可通过网络被访问。4)编写真实的Web服务（Comman Lisp），并将WSMO中的服务描述通过Grounding技术与其正确相连。5)使用扩展的控制的定义进行服务控制描述，并将其嵌入WSMO和IRS系统。

4.4.1 WSMO 构建本体

层次结构匹配服务的知识领域的本体描述主要三个（如图4-11）：1)返回结果的result本体，包括两个子概念：真和假。2)解释本体（如图4-12），包括四个属性，exact（外在），yinxing（隐性），xianxing（显性），score（与节点匹配

分数), 这四个属性的Rang Concepts都是属于Data Type类型中的整性, 且取值个数为无限制。3) 节点列表本体(如图4-13), 包括一个名字为li的属性, 这个属性的Rang Concepts是concept类型中的节点本体, 且取值个数也为无限制。

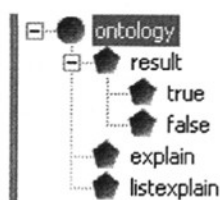


图4-11 层次结构匹配本体

Figure 4-11 Ontology of layer structure match

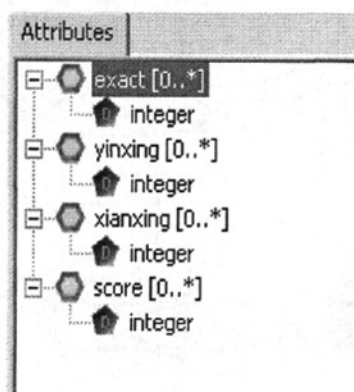


图4-12 解释本体属性

Figure 4-12 Attribute of explain Ontology

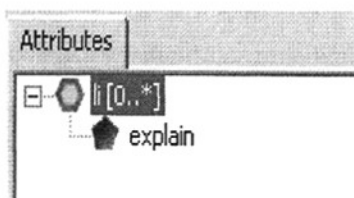


图4-13 节点列表本体属性

Figure 4-13 Attribute of listexplain ontology

4.4.2 WSMO 组件实现及 WSMO 语言描述

由图 3-7 可知, 试卷分析中的层次结构匹配由一个初始化服务和树匹配服务组合而成。而树匹配是在树中的每层进行一次循环的匹配操作, 因此包括一个每层的匹配。以下分开讲述树匹配服务, 层匹配服务及与树匹配链接, 树匹配与层次结构匹配链接。

如图 4-14 所示，树形匹配共包含 7 个 Web 服务描述（是否叶子，叶子返回，扩展节点，层匹配，比较，返回，初始化），相对应具有 7 个目标描述，以及相应的目标描述和 Web 服务描述之间的 7 个 WG-桥服务。服务组合是依靠目标进行抽象流搜索的，因此各目标描述间也存在相应的 GG-桥服务。此 GG-桥服务只存在有前后交互的 Web 服务描述间，如比较目标描述和返回目标描述间，而返回目标描述与初始化目标描述间则不存在这种桥服务，因此共包含 9 个 GG-桥服务。

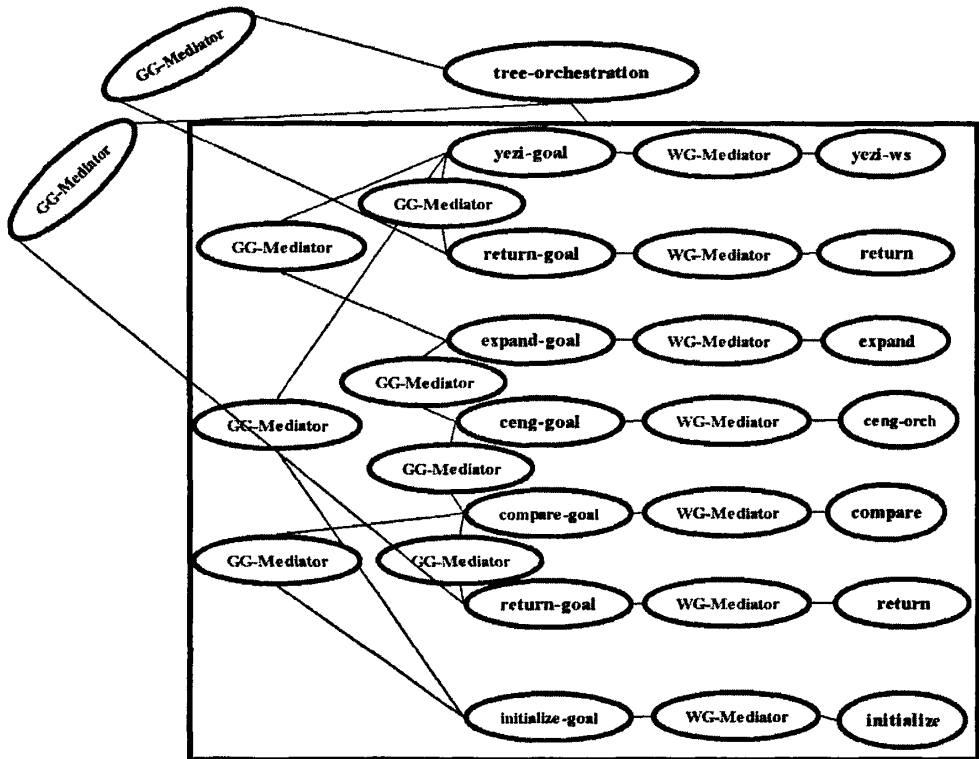


图4-14 树匹配WSMO模型

Figure 4-14 WSMO model of tree match

层匹配的 WSMO 模型如图 4-15 所示。共包含 4 个 Web 服务描述（初始化，返回，取节点，节点匹配），相对应的 4 个目标描述，相应的目标描述和 Web 服务描述之间的 4 个 WG-桥服务，5 个 GG-桥服务。同样道理，返回服务描述与获取节点服务描述间不存在此桥服务，而由于返回服务描述要返回值到层匹配服务，因此它们存在 GG-桥服务。

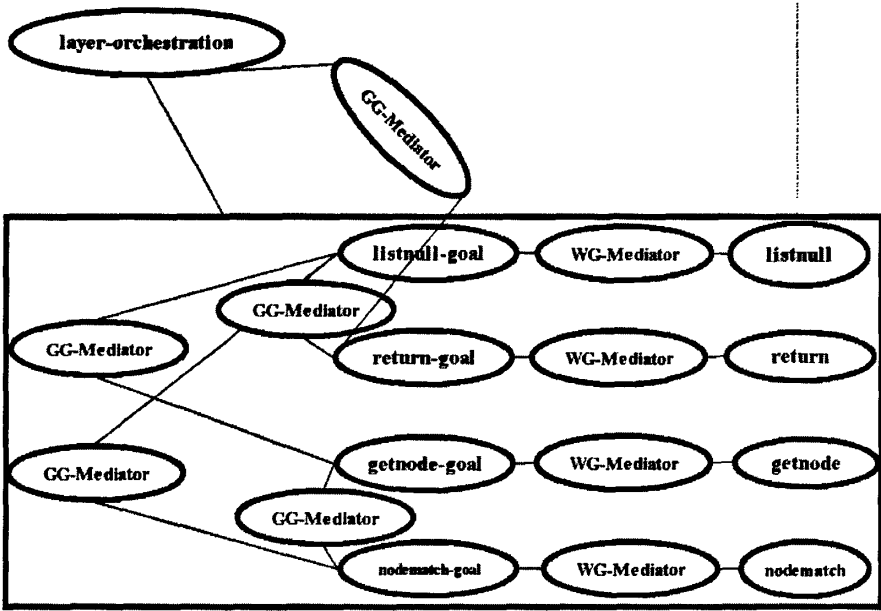


图4-15 层匹配WSMO模型

Figure 4-15 WSMO model of layer match

层次结构匹配服务如图 4-16。树匹配（图 4-14）部分链接于 tree-orch 服务的 orchestration 接口，与初始化服务组合为整个服务。而层匹配组合服务（图 4-15）链接树匹配服务（图 4-14）中的层匹配服务（ceng-orch）的 orchestration 接口。

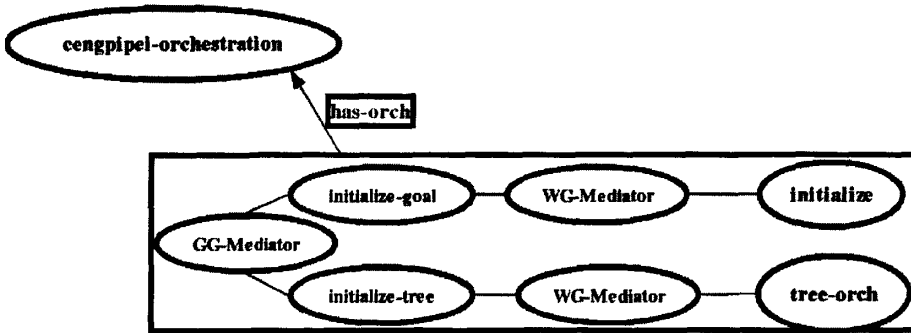


图4-16 层次结构匹配WSMO模型

Figure 4-16 WSMO model of layer structure match

在 WSMO 中构建的各个服务都使用 WSML 语义描述，使用节点匹配服务的 WSML 描述作为例子，节点匹配 Web 服务、节点匹配目标和节点匹配 WG-桥服务的 WSML 语言描述如下。

节点匹配 Web 服务描述：

```
webService nodematch
```

```
importsOntology _ "http://irs.open.ac.uk/cengpipei#ontology"
```

```

usesMediator
    nodematchMediator
interface nodematchInf
    choreography nodematchChor
    stateSignature nodematchState
    importsOntology
        _"http://irs.open.ac.uk/cengpipei#ontology"
    in
        concept explain withGrounding
            _"irslisp://localhost:3001/soap/nodematch",
        concept explain withGrounding
            _"irslisp://localhost:3001/soap/nodematch",
        concept explain withGrounding
            _"irslisp://localhost:3001/soap/nodematch"
    out
        concept result

```

节点匹配目标描述:

```

goal nodematchGoal
    importsOntology _"http://irs.open.ac.uk/cengpipei#ontology"
interface nodematchGoalInf
    choreography nodematchGoalChor
    stateSignature nodematchGoalState
    importsOntology
        _"http://irs.open.ac.uk/cengpipei#ontology"
    in
        concept result
    out
        concept explain,
        concept explain,
        concept explain

```

节点匹配 WG-桥描述:

```

wgMediator nodematchMediator
    importsOntology _"http://irs.open.ac.uk/cengpipei#ontology"
    source nodematchGoal

```

4.4.3 IRS 发布与访问

7 个服务发布到 IRS 中如图 4-17 所示。服务的发布只需将服务描述，目标和桥描述拖动到服务列表中即可，IRS 会将每个服务的名字、本体、输入、输出、服务地址、端口及其使用的本体注册进服务列表。

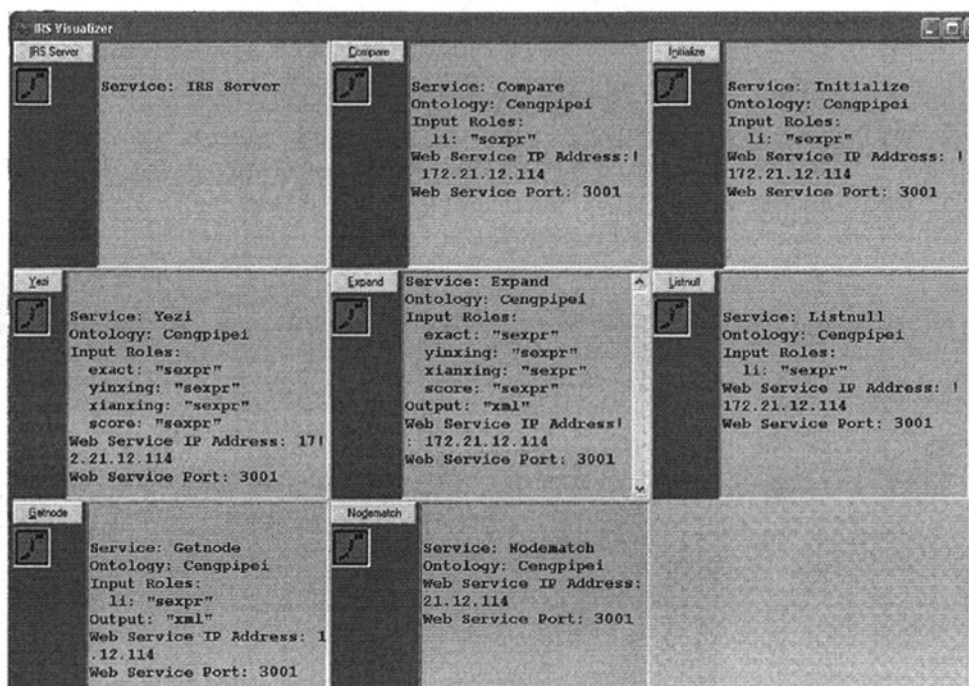


图4-17 层次结构匹配服务发布图

Figure 4-17 publish of layer structure match services

4.4.4 Lisp 服务功能函数

本节讲述 WSMO 中的 Web Service 使用 Grounding 技术与具体的服务链接，和具体服务的 Common Lisp 实现。

在 WSMO 中，以节点匹配服务为例，如图 4-18 所示，它包括一个 Interface，Interface 中有一个 Choreograph 连接，Choreograph 中包含一个 StateSignature。StateSignature 使用 In Node 和 Out Node 模式连接输入和输出。因为节点匹配服务包含三个输入和一个输出，输入包括当前节点，目前最优节点和标准节点，所以 State Signature 包括三个 In Node 输入和一个 Out Node 输出。

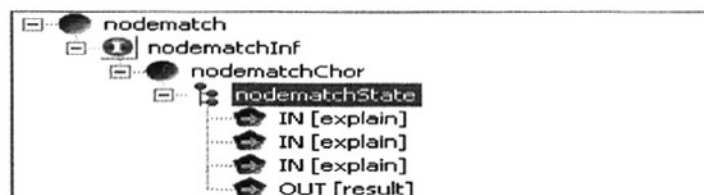


图4-18 节点匹配的StateSignature

Figure 4-18 StateSignature of node match

StateSignature 使用 Grounding 技术与真实服务进行连接，每个 Node 都可以通过添加 Ground 进行与服务的连接，如图 4-19 所示。Localhost 为服务部署的 IRS 平台地址，3001 为端口。三个 In Node 都连接到节点匹配服务。

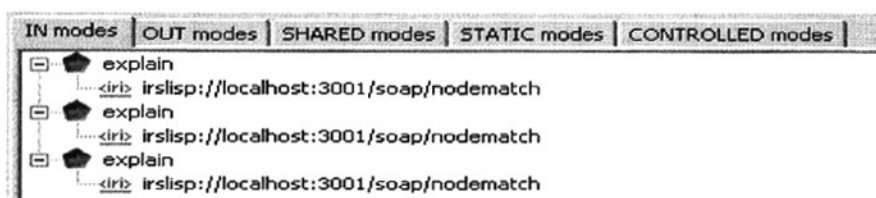


图4-19节点匹配的IN modes

Figure 4-19 IN modes of node match

节点匹配服务使用当前节点和目前最优节点分别跟标准节点进行比较，并将两者的各得的分数进行比较，选出优的节点，如果当前节点优于当前最优节点，则将其赋值给当前节点，并返回。本例子使用的是 LispWorks Personal Edition 5.0.1 平台。

//节点匹配

```
(defun nodematch (opt node care)
  (setq m1 (list-length (set-difference node care)))
  (setq u1 (list-length (set-difference care node)))
  (setq m2 (list-length (set-difference opt care)))
  (setq u2 (list-length (set-difference care opt)))
  (cond ((or (< u1 u2) (and (= u1 u2) (< m1 m2)))
    (setq nownode node)
    (setq result T)
  )
  )
)
```

4.4.5 控制算子的 OCML 语言描述

使用4.2节中对扩展的控制算子的定义和OCML语言描述。将控制关系转换为可执行的部件，以层次结构匹配服务的根节点服务层为例子,服务中的根节点，初始服务，树匹配的定义如下。由于根节点的执行为顺序执行，因此这里还包括层匹配中的顺序执行控制节点。树匹配是循环节点类型，而循环节点由选择，顺序，退出三部分组成，因此还包括中的一个循环控制节点和一个顺序执行控制节点。

层次结构匹配根节点：

```
Cengpipei (structured-orchestration)
    Has-root-component: value sequence1           //包括一个顺
                                                    序执行节点
```

顺序节点 1：

```
Sequence1 (sequence)
    Has-components: value (initialize-ws-1       //包括初始服务
                          Tree-orchestration) //和树匹配
```

初始服务：

```
Initialize-ws-1 (service-component)
    Has-internal-binds: value binding-Tree-orchestration //绑定外
                                                            //在服务
    Has-binding-to-composition-output: value tree-orchestration
                                                            //输出到
                                                            //树匹配
```

树匹配：

```
Tree-orchestration (structured-orchestration)
    Has-root-components:
        value While1           //根节点为 While1
    Has-invocable-description:
        value initialize-goal //可调用前提为初始目标有
```

值

```
    Has-internal-binds:
        value binding-yezi //绑定到叶子处理
```

循环节点 While1 (while)：

```
    Has-ifthen-components:
        value if-yezi-then //包含一个执行判断节点
```


Has-sequence-components:

value Sequence 2 //包括顺序节点 2

Has-return-components: //返回值只来自两个服务节

点

value (Return-yezi-ws
Return-compare-ws)

顺序节点 2:

Sequence2 (Sequence)

Has-components: value (

Expand

Layer-orchestration

If-Compare-then

Initialize2)

4.5 本章小结

本章描述在 WSMO 平台上实现的偏差问题的 PSMs。由于偏差型语义 Web 服务的工程比较大，仅给出其中的顶层组件描述，而选择其中的层次结构匹配部分的项目为例。在层次结构匹配中包括整个偏差型服务建立的所需的五个部分：本体集构建，WSMO 服务的建立，发布到 IRS 平台上，Lisp 功能函数，组合控制算子的扩展语言实现。每个部分都使用节点匹配的代码作为编程事例。

第 5 章 Web 服务自动组合

服务自动组合是指在语义的基础上服务自动进行组合以共同完成一项任务。本章提出一种自动组合（与或组合）方法，并实现组合架构的线性工作流程转换，使用例子进行实验。偏差型语义 Web 服务的自动组合及工作流程是对第三章偏差问题组合架构的进一步扩展，加入了自动组合和工作流的研究。

5.1 传统规划方法

目前,传统的解决服务组合的问题包括: workflows 组合 (如 Eflow 2000, Benatallah 2002)、动态地自动化规划、基于程序综合的组合方法。基于 workflow 的方法分为两类: 静态 workflow、动态产生 workflow。动态地自动化产生规划的方法包括: 基于情境演算的规划、基于规划域定义语言的规划 (如 PDDL 2002)、基于规则的规划 (MEDJAHED 2003, SWORD 2002)。规划方法假设每一个服务使用所规划情境下的前提条件和后驱状态, 按照某种规划方法选择执行服务改变世界的状态^[37]。

5.2 与或图 Web 自动组合

本章提出一种使用与或图解决 Web 服务自动组合问题的方法。此方法将任务分解为更小的与、或子任务, 使用相似性算法匹配服务并使用启发式算法搜索以找到最优解图, 嵌入的 SEQUENCE-AND 类型节点支持从解图到执行 workflow 的转换, INTERFACE 类型节点支持 workflow 中的用户交互和正确性检验, 并使用实例进行实验说明。

5.2.1 引言

服务可以表述为一个行为, 因此服务自动组合问题可以利用人工智能规划技术进行处理^[38]。如[39]所提出的, 在传统的人工智能中, 与或图搜索是处理空间搜索问题的一个重要的工具。因此, 与或图搜索技术可以用于处理 Web 服务组合问题, 它支持将任务分解成较小的与、或子任务直到所有的任务是可解的, 对应服务由下一层子任务或叶子节点组成。执行此操作过程的空间搜索得到一个使初始状态满足的解图, 并根据其生成一个可执行线性 workflow。

5.2.2 组合实例

我们以一个旅行领域的情景作为例子。客户的需求是为一次旅行预订一个航班和酒店，且必须满足：首先，找到可用的航班和酒店；其次，客户要亲自选定合适的特定航班；最后，酒店的日期应该与飞机到达的日期一样，即酒店的预订应等到飞行的预订后。而且还存在一个情况就是提供航班和酒店服务的 Web 服务有很多。

5.2.3 与或图组合

5.2.3.1 相似度计算

服务的输入集描述一个服务期望从用户得到的值，而服务的输出集描述一旦输入值满足时服务所传递出的值。假定服务的输入和输出使用 W_i 和 W_o 表示，而用户的输入和输出为 R_i, R_o 。基于概念相似性的服务可连接性即指输入与输出满足一定的条件，有两层的含义：第一层表示基本的可连接性，即第一条条件 (First Condition): $R_o \subseteq W_o$ ，即 Web 服务可以提供用户所需的值作为输出； $W_i \subseteq R_i$ ，即用户可以提供所有服务必须的输入值。第二层讨论输入与输出集的相似度的大小（表示为集合 A 和 B），即第二条条件 (Second Condition): 首先，从集合 A 中取元素 A_i ，然后将 A_i 与 B 集合每一个元素进行匹配，从中得到一个与 A_i 具有最大相似度的元素的相似度作为 A_i 与 B 的相似度，最后计算所有 A_i 的相似度的平均值作为 A 与 B 的相似度。在输出集匹配中，A 代表 R_o 而 B 代表 W_o 。相反，在输入集的匹配中，B 代表 R_i ，而 A 代表 W_i 。

5.2.3.2 组合服务的与或图形成

与或图中的节点表示一个可运行 Web 服务的封装实体，它包括状态和一系列行为。图中的有向边表示输出一方的输出集满足输入一方的输入集，即子集。边上的权重由函数 $\delta (S_1, S_2)$ 根据相似条件的第二条条件定义，表示两个服务间的距离以支持 AO* 的启发式搜索。我们定义函数 δ 的取值范围在 0 和 1 之间，并且 0 表示相关度最低，而 1 意味着相似性最高。与或图包括：包含最终的输出的反映客户需求的 START 根节点，子服务都可解才能可解的 AND 节点，任何一个子服务可解就可解的 OR 节点， δ 值为 0 的不可解的 NONTERMINAL（非终端）节点，包含用户输入的 TERMINAL（终端）节点。使用 AO* 启发式算法搜索最优解图 $G(S)$ ， $G(S)$ 对应于以 S 作为 START 节

点的解方案。G (S) 具有以下属性:

1) 在 G 中, 如果 n 是一个 OR 节点并且 n 在 G (s) 中, 则至少有一个 n 的直接子节点在 G (s) 中。

2) 在 G 中, 如果 n 是一个 AND 节点并且 n 在 G (s) 中, 则所有 n 的直接子节点在 G (s) 中。

3) G (s) 中的任何每个端节点都是 TERMINAL 的子集

4) 不存在 NONTERMINAL 节点。一旦 NONTERMINAL 节点被扩展, 就不可能进一步推理而相应的子解图将被放弃。

5) 不存在除 S 和 S 的子集点外的其它节点在 G (s) 中

现在, 我们定义一个状态空间 (原子服务和组合服务), 原子服务, 开始状态, 目标状态和推理规则 (组合操作)。解图搜索过程如下:

开始状态:

- 一系列具有不同状态空间的 Web 服务作为本地可用节点;
- 一个单一的 START 节点, 包含组合服务的输出, 代表用户需求;
- 一系列有限 TERMINAL 节点, 包含组合服务的输入, 代表目前可用的用户输入;

结果: 一个优化的组合服务图作为解

搜索执行:

搜索算法执行以下步骤:

1) 传统 AO*启发式算法: 它包括两个主要操作

- 自上而下的扩展操作, 以寻找潜在解图包括基于第一条件的有向边的可连接性和基于第二条件的相似性评价权重, 直到服务的输入是一个存在 TERMINAL 节点的子集。如果权重值低于某特定值或节点是 NONTERMINAL 节点, 则此节点应不被扩展。
- 自下而上的成本 (路径权重) 修改操作, 从而更新节点的权重值, 以找到具有低成本的最优解图。

2) 转换可执行 workflow: 一个搜索状态代表了一系列可执行步骤, 此步骤从 TERMINAL 节点开始到 START 节点并满足服务组合要求。从解图转换到可执行 workflow 的执行将在下一节进行详细讲述。

5.2.4 转换成可执行 workflow

5.2.4.1 转换 workflow 实例

这里使用[40]中的一个典型例子。假设目前有服务 S1 和 S2 : 输入 (S1) =

(A, B), 输出 (S1) = (C) , 输入 (S2) = (B, C) 中, 输出 (S2) = (D)。无论 S1 和 S2 都不能满足请求服务 R: 输入 (R) = (A, B), 输出 (R) = (C, D)。两个服务组合的以下工作流程才能解决此任务: S1 服务被调用, 输入为 (A, B) 并产生输出 C, 然后 (A, B, C) 可用, 且能满足 S2 的输入需求, S2 服务能产生输出 D。

5.2.4.2 SEQUENCE-AND 节点

与或图中加入 SEQUENCE-AND 节点是个规划过程。在扩展的 AO*启发式算法中, 规划器自上到下使用深度优先扩展图形, 直到服务的输入是可用 TERMINAL 节点的子集。其次, 规划器自下到上收集服务的输出, 直到它遇到非可用节点。第三, 规划器继续扩展 AND 节点的其余子节点, 直到开始节点。

图 5-1 是将 SEQUENCE-AND 节点嵌入到以上 AO*启发式算法的一个例子。我们假定目前规划器已经使用深度优先算法扩展图到 D1 和 D2 节点, 而且发现 D1 和 D2 节点的输入都可由 TERMINAL 节点满足, 然后它收集 D1 和 D2 节点的输出并往上返回, 并标注 C, B 节点为可用节点, 收集 C, B 节点的输出, 直到 START 节点的 E 节点。规划器在 E 节点处理中, 从 TERMINAL 节点, B, D2 节点中找到输入的满足条件, 则标记 START 或节点为 SEQUENCE-AND 节点类型。

SEQUENCE-AND 本身代表从左至右的默认执行顺序序列。SEQUENCE-AND 节点的执行 workflow 为: 执行器在遇到 SEQUENCE-AND 节点后将等待直到右边的节点等待到左边的输入。具体的示例入图 5-1 所示, 执行器并行调用 D1, D2, E 节点。但对于 E 节点而言, 执行器将挂起 E 节点, 直到左边输入的触发唤醒。它首先执行并同时收集 D1, D2, 其次 C 和 B 节点, 然后触发唤醒 E, 直到 START 节点变为可用节点。

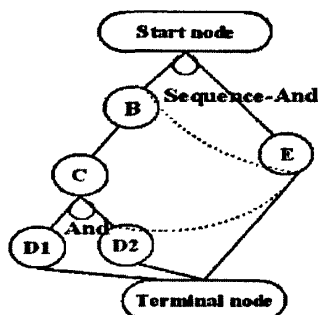


图5-1 SEQUENCE-AND节点例子

Figure 5-1 Illustration of SEQUENCE-AND node

图 5-2 为 5.2.4.1 中的例子的图形化的表示，它能满足 R 需求的工作流。

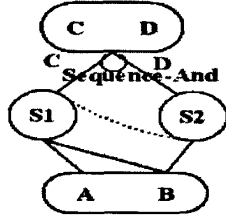


图5-2例子（5.2.4.1）执行工作流

Figure 5-2 Workflow illustration of example (5.2.4.1)

5.2.4.3 执行构造

规划组合的执行集合包括：Sequence, Unordered, Choice, Switch, If-Then, Split and Split+Join^[41]。我们定义可执行节点有四种类型：AND 节点，OR 节点，SEQUENCE-AND 节点，SWITCH-OR 节点。AND 节点和 OR 节点的定义与前面与或图中的定义一样。SEQUENCE-AND 节点是 AND 节点的一种，使用 \oplus 表示。SWITCH-OR 节点是 OR 节点的一种，使用 \odot 表示。每一个执行节点代表一种执行步骤。图 5-3 为执行节点与对应的执行集合组合的执行结构的图形表示：

- AND 节点表示 Split 执行，即从左到右顺序执行。边表示满足所需的输入集是其它服务输出集的子集。
- SEQUENCE-AND 节点表示 Split+Join 执行。
- OR 节点表示 Choice 执行。OR 节点的子节点从左到右是同等的。
- SWITCH-OR 节点表示 Switch 执行，表示只能选择一个子节点。
- 路径表示自下而上序列 Sequence 执行。
- 不同的路径默认表示无序，除非遇到 SEQUENCE-AND 节点。
- 图的边表示 If-Then 条件执行，即所需的输入集是其他服务的输出集的子集。当输出为空时（NULL）时，执行器将停止，例如今天没有用户需要的飞机航班搜索结果。

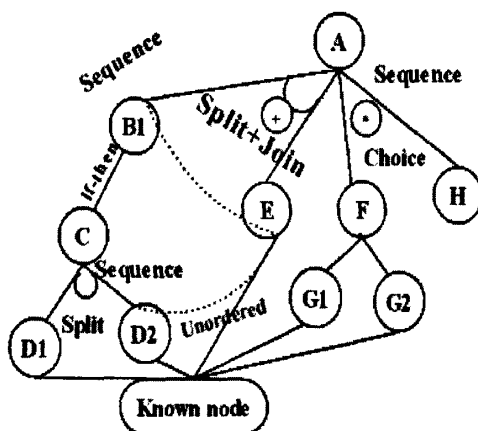


图5-3 执行结构

Figure 5-3 Execution Constructs

5.2.4.4 INTERFACE 节点

在组合和执行过程中，用户应能够动态监测整个系统，也包括用户偏好的选择。目前，INTERFACE 节点分为两种类型：USER PREFERENCE 节点和 CHECK 节点。INTERFACE 节点的特点为：指定 INTERFACE 节点的权重为 0 代表无影响；在 INTERFACE 节点满足的条件下，规划器应首先考虑扩展 INTERFACE 节点。

CHECK 节点可以作为变量类型和一致性检查^[42]。CHECK 节点的设计规则是具有相同的期望检查值作为输入和输出。例如，我们可以设定一个 CHECK 节点具有相同的飞机票值的输入和输出，以检查此飞机航班是否是用户期望的航班，然后决定是否调用银行缴费服务。CHECK 节点的执行规则是：如果 CHECK 节点失败，执行器将停止。

USER PREFERENCE 节点的设计为所需作出判断的请求作为输入，用户偏好作为输出。例如，USER PREFERENCE 节点包含今天的飞机列表作为输入，一个航班选择作为输出。USER PREFERENCE 节点的执行规则是：如果输出为空，执行器将停止。

5.2.5 组合实例的可执行 workflow

这一节将通过旅行例子对服务的自动构建和转换为 workflow 进行实验，如图 5-4 所示。

开始状态：

- 组合 Web 服务的所需输出，即飞机票和房间号是组合 Web 服务的

START 节点的输出。

- 组合 Web 服务的已知输入，即飞行日期和目的地是组合 Web 服务的 TERMINAL 节点的输出。
- 一个 USER PREFERENCE 服务节点，输入为航班列表，输出为一个航班编号。

搜索结果：一个解图，如图 5-4 所示。

规划执行：

- 首先，规划器发现名为 Book Ticket 的 Web 服务，它具有 Ticket 的输出，和一个名为 Book Hotel 的 Web 服务，它具有 Room ID 的输出。
- 规划器扩展 START 节点到 Book Ticket 节点和 Book Hotel 节点，并标记 START 节点为 And 节点。
- 然后，基于 Flight ID 值，一个 USER PREFERENCE 节点被发现与 Book Ticket 连接，并且由于 Book Hotel 节点需要到达日期的输入 START 节点被标记为 SEQUENCE-AND 节点。
- 一个名为 Flight Search 的 Web 服务可以输出航班列表且输入（日期，目的地）为 TERMINAL 节点输出集的子集，规划器则标记其为可用节点，并认为此路径是可通的。
- 从下至上，规划器收集每个 Web 服务的输出，特别是 Book Ticket 服务的输出（到达日期）
- 由于目标日期的可用，Book Hotel 服务被唤醒并标记为可用节点。规划器收集 Book Hotel 服务的输出并标记 START 节点为可用节点。

组合任务是一个满足组合需求的任务集的工作流。执行程序由具体的输入值激活。执行步骤如下：Flight Search 服务得到日期和目的地的输入则输出航班列表，同时 Book Hotel 服务获得目的地的输入并以 SEQUENCE-AND 节点模式等待到达日期值的唤醒，用户在 Client Interface 节点中从航班列表选择一个航班，Book Flight 服务获得航班的输入并输出到达日期。同时，执行器收集各服务的输出，包括目的地输入，激活了 Book Hotel 的 SEQUENCE-AND 节点。Book Hotel 输出房间号则到达 START 服务，执行完毕。

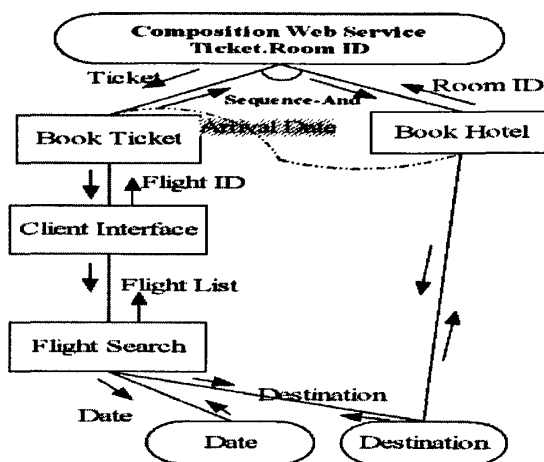


图5-4 旅行代理 workflow

Figure 5-4 Workflow of Travel Agency

5.3 嵌入 WSMO 的与或组合体系

大多数的服务自动组合方法都采用特定的自动化组合系统来组合服务,利用规划系统直接产生组合服务,而都没有考虑语义转化及异构消除问题。OWL-S 就是一个与智能规划联系紧密的服务语言,而 WSMO 研究的组合技术是基于知识工程角度的,其中的桥组件使用数据绑定技术处理各种异构的问题,而且 WSMO 框架在语义层上区分目标(要实现的任务的抽象定义)和 Web 服务(可以实现目标的服务的描述),即可以在服务和目标之间建立可变的映射,从而允许多对多的动态的问题与解决方法映射和基于知识发现的服务选择,因此可以支持功能驱动的服务调用。

与或图自动组合方法嵌入到 WSMO 模型中,体系结构如图 5-5 所示。此体系结构主要分为四层,从上到下依次为知识本体层(Knowledge Ontology),任务模板层(Task template),桥连层(Mediator),Web 服务层(Web Services)。知识本体层用于存储领域知识的本体表示,以支持任务模板层的任务间的输入与输出的相似度计算;任务模板层由任务组件组成。此处的任务组件与 WSMO 中的目标组件类似,不同处为目标描述用户的需求,而任务用于描述实现一个服务的自动组合所必须的信息,包括此任务被加入组合前和加入后的规则指定。任务存放于任务数据库(RDBMS)中,组合规划器(Planner)在 RDBMS 中执行任务模板的搜索以用于与或服务的组合与执行工作流转换;任务组合解图确定后,使用语义描述定位与此匹配的 Web 服务,有两种情况存在:任务与 Web 服务为一对一关系,一个 Web 的组合信息描述为一个任务直接放于任务库中;任务与 Web 服务为一对多关系,一个 Web 的组合信息描述与库中的

一个已有的任务建立关联；任务模版的一次具体的输入值触发一次具体工作流的形成，并作为输入来调用相应的底层的 Web 服务，桥用于解决任务与 Web 服务的异构问题。

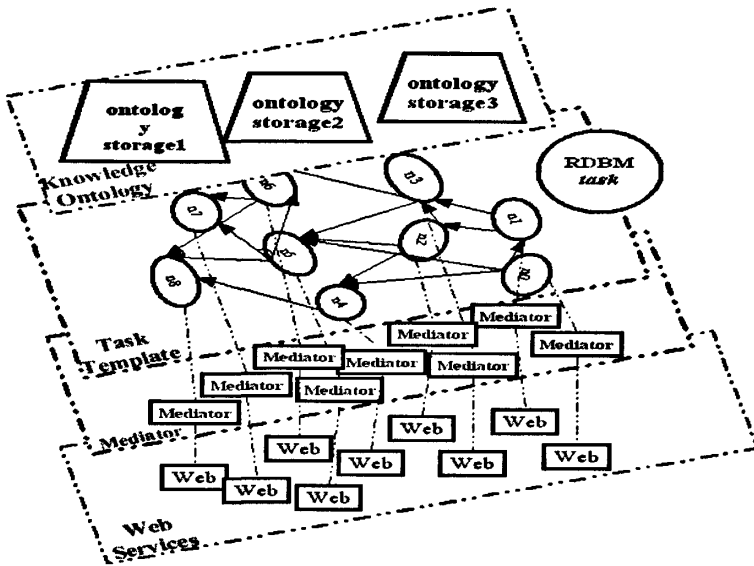


图5-5 与或服务组合体系

Figure 5-5 Web Composition Architecture of And/OR

体系的执行步骤为：首先，基于 Web 服务的输入和输出构建任务模板。第二，执行自动组合和可执行工作流的转换。第三，使用能力-驱动来匹配 Web 服务与任务模板，并使用桥进行异构映射。

5.4 本章小结

单个简单的 Web 服务往往无法满足实际的应用需求，人们常常需要组合多个 Web 服务以完成一个较复杂的任务，这就是服务的组合问题，而服务的自动组合更是以后的发展方向。同样，偏差型语义 Web 的自动组合可促进偏差问题的求解和知识组件的重用。因此，为了促进偏差型语义 Web 服务的发展，本章提出了一个初步的使用与或图的服务自动组合方法，并加入了自定义的介于与或关系的节点间关系类型，实现与或解图到可执行的工作流的转换。

第6章 语义 Web 服务的知识处理平台的讨论

基于知识的系统分为两部分：分析型和综合型，其中，分析型系统一般包括分类、诊断、评估、监测等类型，综合型系统包括设计、规划、建模、调度分配等类型。基于分析型的偏差问题的语义Web服务称为偏差型语义Web服务，综合型系统对应的语义Web服务本文称为解决型语义Web服务，知识型语义Web服务包括偏差型语义Web服务和解决型语义Web服务。课题组研究的最终目标是构建一个可处理知识型语义Web服务的知识处理平台，本文称之为语义Web服务的知识处理平台，凡是通过其PSMs可转换纳入知识处理平台的Web服务组合架构中的系统可称之为知识型系统。由于解决型语义Web服务和偏差型语义Web服务的区别主要是Web服务的组合架构不同，而对于与知识处理平台有关的理论问题则是相同的。例如：服务的自动组合、Web服务的粒度大小、将Web服务的组合转换成可执行的工作流等问题。因此，在下文中，本文都是基于前述的关于偏差型语义Web服务的工作来进行对知识型语义Web服务的知识处理平台的讨论。

语义 Web 服务的知识处理平台依据知识型系统的语义 Web 服务的组合架构在一系列的知识型语义 Web 服务中自动查找与组合，并将得到的 Web 服务组合转换成可执行的工作流加以处理。以上是从 PSMs 角度给出了知识处理平台的粗略思想，而如果具体到知识处理平台与知识型语义 Web 服务的实现，则采用本体描述领域知识，任务组件描述一个服务的自动组合所必须的信息，Web 服务组件语义的描述服务的语义功能，桥组件执行数据绑定处理异构问题，整体体系结构如图 5-5。

由于知识处理平台的复杂性，本文只对于其中的几个关键问题：语义 Web 服务的自动组合与工作流，知识级语义 Web 服务集合，原子级语义 Web 服务的粒度大小与完备集合，Web 服务语义功能描述进行了讨论。

6.1 语义 Web 服务的自动组合与工作流

知识平台的服务自动组合限制在知识型系统的语义 Web 服务的组合架构中，具有一定的范围限制，有助于自动组合方法的实现。以偏差问题的语义 Web 服务组合架构为例，最高层由三个部分组合而成：偏差数据，关系对应和解释偏差，其中的偏差数据包括：观测抽象、标准抽象、偏差抽象。三部分在树状组合图中为与的关系，因此转换到可执行工作流中应为顺序执行，如图 6-1 所示。在第三章中已经提出，不同的领域问题对应偏差问题的语义 Web 服务组合架构的不同实例化，如水污染系统与试卷分析系统，因此对应的线性工

作流的组成部分也是不同的。知识处理平台中，针对特定领域，服务组合中的三部分如何自动地判断和组合将是一个很大的挑战。例如，观测抽象可能直接连接到关系对应，而跳过标准抽象与偏差抽象。而且服务间可能通过不同的路径相连，如观测抽象可能通过不同的路径连接到标准抽象。一个足够好的搜索算法对于不同路径的选择是有很大帮助的。

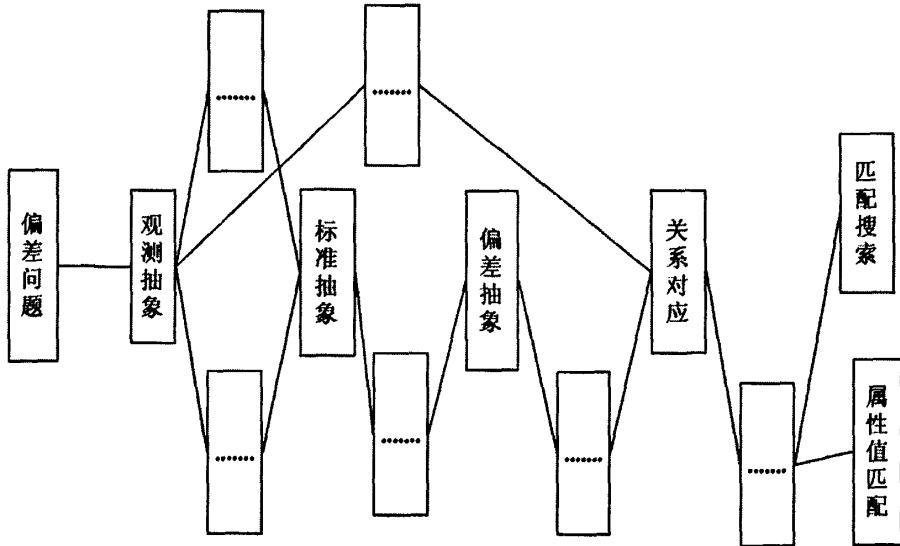


图6-1 偏差问题Web服务线性组合

Figure6-1 Line composition of Web services of deviation problem

在知识型系统的语义Web服务的组合架构中，Web服务间同样为与、或关系，因此知识型系统的语义Web服务的组合架构也同样可以采用与或图组合方法。即知识处理平台中可以使用与或图表示服务的组合，并使用扩展的启发式算法搜索解图，并使用工作流转换方法将与或解图转换为可执行的工作流。对于知识处理平台中的服务组合，与或图组合方法仅是课题组目前给出的一个初步方法，关于其应用细节与更适合的组合方法的探索等尚需进一步的研究。

6.2 知识级语义 Web 服务

知识系统中处理问题的方法有很多种，最常用是推理。推理机中的推理方法与控制策略决定了专家系统对知识库中知识的有效应用。组合架构中的知识级语义Web服务有多种类型，产生这个问题有多方面的原因：1) 推理方法有多种类型，包括演绎推理，归纳推理，默认推理，不确定推理，单调推理，非单调推理。2) 推理策略包括推理方向，搜索策略，冲突消解策略，求解策略，限制策略。3) 从推理方向上又可分为正向推理，反向推理，正反向混合推理。4) 知识有两种类型：一种是事实或普遍的规律；另一种是具有启发性的或试探性

的知识，它们是专家经过多年工作凭经验得到的知识，适于使用具有启发功能的方法。5) 知识表达方式也有许多种，如产生式系统，框架，语义网络，概念结构等，对于表达不同的知识具有各自的优势。

知识级语义Web服务有如此多的类型，而语义Web服务组合架构工作流的形成即在这些知识级语义Web服务中自动查找，组合并转换成为可执行工作流。因此，在知识级语义Web服务方面，架构的可执行工作流的自动构建存在两个难题：1) 知识级语义Web服务集合应该包括以上情况的所有组合，这将是一个庞大且复杂的系统，而且通过不断的服务组合的学习和经验的积累，将会有新的组合出现用于解决某一类特定问题。因此，如何构建一个比较适宜的知识级语义Web服务集合尚需进一步的研究。2) 选择不同的知识表达方式对问题解决的难易程度有很大的影响。例如同样可以使用正向推理的情况下，选择不同的控制策略对问题解决有很大影响。多年的人工智能专家的经验告诉我们，每个问题通常都会对应一个相对适宜的解决方法。因此，知识处理平台中的自动组合执行器能否正确的选择一个最适宜的知识级语义Web服务对应解决一个任务问题也仍需进一步研究。

6.3 原子级语义 Web 服务

知识型语义Web服务由知识级语义Web服务与原子级语义Web服务组合构成。原子级语义Web服务是一定粒度的具备一完整功能的服务。为了提高知识服务的可重用性，需要制定合理并完备的原子级的语义Web服务集合，而且原子级的语义Web服务的粒度的大小对于集合的制定和可重用性的有效利用也有很大的影响。因此，在原子级语义Web服务方面，存在两个难题：1) 如何建立完备的原子级的语义Web服务集合。2) 如何制定架构的终端节点-原子级语义Web服务的粒度大小。

6.4 Web 服务语义功能描述

知识型语义Web服务的功能性描述包括两个方面的内容：1) 更精确的定义知识级语义Web服务间的匹配度定义，如在服务的输入和输出中加入语义化规则，而不仅仅局限于输入与输出集合的包含关系。2) 加强服务功能的语义描述，促进较优Web服务的查找，如根据A服务的输出功能描述，A的输出知识具有启发性特点，查找到B服务，原因在于B服务的功能描述显示此服务具有处理启发式知识的偏好。

6.5 本章小结

目前对于知识处理平台的研究只是完成了粗略的结构思想和基础理论研究, 包括提出了 PSMs 及其对应的组合架构, 知识型系统的语义 Web 服务的构建, 组合控制算子, 自动组合方法, 及组合方法嵌入到 WSMO 的体系结构。还有很多细节方面需进一步研究: 将与或图组合方法应用于知识型语义 Web 服务的组合和工作流的转换尚有不完善之处; 由于其服务的多样性特点, 知识级语义 Web 服务和原子级语义 Web 的集合构建仍是一个很大的挑战, 需要进行大量的实验研究和证明。

结论

基于知识的系统可分为两部分：分析型和综合型，其中，分析型系统一般包括分类、诊断、评估、监测等类型，综合型系统包括设计、规划、建模、调度分配等类型，但这种传统的分类方法并没有给出一个统一的处理模型。本文提出偏差问题并给出了相应的偏差问题PSMs，并给出了与PSMs对应的偏差问题的服务组合架构，将偏差问题的求解转化为语义Web服务的组合。然后对组合架构的一系列支持方法进行了研究，包括在组合架构中加入组合控制算子控制服务间执行关系，与或图组合方法支持解图的搜索与转换为可执行 workflows，并给出了与或组合方法嵌入到WSMO中的体系结构。最后对于基于知识型系统的语义Web服务组合架构的知识处理平台给出了粗略思想和关键问题的讨论。其中有一个部分为组合架构的语义Web服务描述实现。本文的任务是对于分析类型的问题进行研究，而关于综合类型的问题的研究由另一位同学完成。

本文中，研究工作按章节顺序依次讲解：第二章提出偏差问题概念并给出了相应的PSMs；在第三章中，对偏差问题的PSMs的概念间的关系进行讨论，提出了偏差问题的服务组合架构。为支持服务间执行关系在架构中的正确反映和执行控制的实现，扩展组合架构加入了控制算子；在第四章中，由于WSMO是构建知识系统的语义Web服务的平台，因此选择WSMO平台实现了组合架构的语义Web服务描述，通过IRS进行发布。嵌入OCML语言实现的服务组合控制算子，并在实例中给出了具体实现；在第五章中，设计使用与或方法自动组合服务并转换为可执行的线型工作流，即讨论如何将分层树状结构转换为线型的可执行工作流，并给出嵌入WSMO的组合体系；第六章对语义Web服务的知识处理平台进行了初步讨论。

本文的创新如下：

- (1) 根据偏差问题中领域知识的多少，将知识工程中的分析型系统加以扩展后定义为偏差问题，并提出了相应的PSMs。
- (2) 根据PSMs提出偏差问题的语义Web服务组合架构。组合架构中的服务由一定粒度的知识级语义Web服务和原子级语义Web服务组合而成，呈分层树状结构。设计并加入了控制算子，支持服务间执行关系在架构中的正确反映与组合服务向可执行过程的转换。
- (3) 基于WSMO体系，给出了一个使用与或图的自动服务组合方法，及将与或解图转换为可执行的工作流，以支持组合架构中的分层树状组合图的实现和其到线性可执行性工作流的转换，并给出了对语义Web服务组合架构进行处理的体系结构。

(4) 针对构建语义 Web 服务的知识处理平台的目标, 给出了平台构建的粗略思想和几个关键问题的讨论。

对偏差型语义 Web 服务的研究还有很多需要进一步深入研究的地方。我们的最终目标是希望构建一个语义 Web 服务的知识处理平台, 抽象并设计出一定粒度的原子级语义 Web 服务, 提供服务间的自动发现与组合, 从而实现知识处理构件的可重用性。本文在最后一章仅是给出了一个粗略的思想, 而且由于知识领域的复杂性以及实验数据不是很充分, 最终目标的实现还有许多细节问题有待研究。下述的几个方面的工作有待于进一步研究和实现:

- (1) 认真分析各类 PSMs 中可重用的语义 Web 服务, 提炼出比较完备的一定粒度的知识级、原子级语义 Web 服务。
- (2) 对偏差型服务的语义功能描述进行研究, 涉及输入与输出的功能描述和规则定义, 服务的功能语义描述, 目标与服务的语义匹配。
- (3) 在偏差问题的组合架构中, 不同的知识系统对应不同的最终架构, 如何实现服务自动识别和组合下一个组件还是一个很大的挑战。

参考文献

- 1 Motta,E, Domingue,J, Cabral, and Gaspari.M. IRS II: A Framework and Infrastructure for Semantic Web Services[J]. 2nd International Semantic Web Conference.
- 2 Domingue, J., Cabral, L., Galizia, S., Tanasescu, V., Gugliotta, A., Norton, B., and Pedrinaci, C. IRS-III: A Broker-based Approach to Semantic Web Services, *Journal of Web Semantics*, 6, 2, pp. 109-132, Elsevier, 2008.
- 3 Domingue. J, Cabral, L. Hakimpour.F, Sell.D. And Motta, E. IRS III: A Platform and Infrastructure for Creating WSMO based Semantic Web Services, *Workshop: WSMO Implementation Workshop*.
- 4 Hakimpour, F., Sell, D., Cabral, L., Domingue, J., and Motta, E. Semantic Web Service Composition in IRS III: The Structured Approach [J]. *International IEEE Conference on E Commerce Technology*, Germany, 2005.
- 5 Stollberg, M., and Norton, B. A Refined Goal Model for Semantic Web Services [J]. *The Second International Conference on Internet and Web Applications and Services*, Mauritius, 2007.
- 6 Viceconti, M., Taddei, F., Jan, S., Leardini, A., Clapworthy, G., Domingue, J., Galizia, S., Quadrani, P. *Bioengineering Modeling and Computer Simulation [J]* . Y. Gonzez and M. Cerrolaza (Eds.) , *CIMNE*, Barcelona, Spain, 2007.
- 7 Lambert, D. Integrating services for integrated medicine [J]. *Workshop: Semantic Web in Ubiquitous Healthcare at 6th International Semantic Web Conference*, Busan, South Korea, 2007.
- 8 Tanasescu, V., Gugliotta, A., Domingue, J., Villarias, L., Davies, R., Rowlatt, M., Richardson, M., and Stincic, S. *Geospatial Data Integration with Semantic Web Services: the eMerges Approach [J]*. *The Geospatial Web*, eds. Arno Scharl, Klaus Tochtermann, Springer, 2007.
- 9 Gugliotta, A., Tanasescu, V., Domingue, J., Davies, R., Gutiérrez-Villarias, L., Rowlatt, M., Richardson, M., and Stini, S. *Benefits and Challenges of Applying Semantic Web Services in the e-Government Domain [J]*. *Austrian Computing Society (OCG)* , 2007.
- 10 Hakimpour, F., Domingue, J., Motta, E., Cabral, L., and Lei, Y. *Integration of OWL-S into IRS-III*. *Workshop: AKT Workshop on Semantic Web Services*, KMi, The Open University, Milton Keynes, UK, 2004.
- 11 Cabral. L, and Domingue. J. *Mediation of Semantic Web Services in IRS-III*, *Workshop: Mediation in Semantic Web Services (MEDIATE 2005)* at *International Conference on*

- Service Oriented Computing (ICSOC 2005) .
- 12 史忠植. 知识发现. 北京: 清华大学出版社, 2002.
 - 13 梁晟. 基于语义 Web 的服务自动组合技术的研究[D]. 中国科学院研究生院博士学位论文, 2004,5.
 - 14 Zhizhong.Shi, Yongquan.Liang, Bin.Wu. Knowledge Engineering and Management [M]. Machinery Industry Press.2003.
 - 15 William J. Clancey. Heuristic Classification. Artificial Intelligence, 1985, 27(3): 289-350.
 - 16 Richard Benjamins and Dieter Fense. Editorial: problem-solving methods[J]. International Journal of Human Computer Studies, 1998, 49 (4) : 305-313.
 - 17 Enrico Motta, Wenjin Lu. A Library of Components for Classification Problem Solving.2004.
 - 18 马玉莲. 面向解释的偏差分析的模型化[D]. 北京工业大学工学硕士学位论文, 2007.
 - 19 王鑫, 王宇冬, 马玉莲. 面向解释的偏差分析问题解决方法[J]. 计算机研究与发展, 2007.
 - 20 Dieter Fensel, Enrico Motta, V. Richard Benjamins, et al.. The Unified Problem-solving Method Development Language UPML. Knowledge and Information Systems, 2003, 5(1): 83-131.
 - 21 Dave Lambert , Enrico Motta . Operational Concept Modelling Language [J].<http://kmi.ope.ac.uk/projects/ocml/>.
 - 22 John Domingue, Enrico Motta, Oscar Corcho Garcia. Knowledge Modelling in WebOnto and OCML[J]. A User Guide. <http://kmi.open.ac.uk/projects/ocml/>
 - 23 Domingue, J., Cabral, L., Hakimpour, F., Sell, D., and Motta, E. (2004) IRS III: A Platform and Infrastructure for Creating WSMO based Semantic Web Services. Workshop: WSMO Implementation Workshop, Frankfurt, Germany, 2004.
 - 24 Stollberg. M, Moran. M, Cabral. L., Norton. B, and Domingue.J. Experiences from Semantic Web Service Tutorials, Workshop: Semantic Web Education and Training Workshop (SWET'06) at First Asian Semantic Web Conference (ASWC) .
 - 25 Hakimpour, F., Sell, D., Cabral, L., Domingue, J., and Motta, E. Semantic Web Service Composition in IRS III: The Structured Approach. International IEEE Conference on E Commerce Technology, Germany, 2005.
 - 26 Jacek Kopecky , Dumitru Roman , James Scicluna. WSMO Use Case: Amazon E-commerce[J].<http://www.wsmo.org/TR/>.
 - 27 Jos de Bruijn, Holger Lausen, Reto Krummenacher, Axel Polleres, Livia Predoiu, Michael Kifer , Dieter Fense . The Web Service Modeling Language WSML[J]. <http://www.wsmo.org/TR/>.

- 28 Kerrigan. WSMO Editor[M]. <http://www.wsmo.org/TR/d9/d9.2/v0.3/20050613/>
- 29 Michael Stollberg,Rubén Lara. WSMO Use Case "Virtual Travel Agency"[J]. <http://www.wsmo.org/TR/>.
- 30 Marin Dimitrov , Vassil Momtchev , Alex Simov , Damyan Ognyanoff. wsmo4j Programmers Guide.<http://www.wsmo.org/>.
- 31 Jacek Kepecky , Matthew Moran,Tomas Vitvar,Dumitru Roman. WSMO Grounding <http://www.wsmo.org/>.
- 32 Galizia. S and Domingue. J. Towards a Choreography for IRS-III[J]. Proceedings of the Workshop on WSMO Implementations (WIW 2004)
- 33 Dumitru Roman, James Scicluna, Cristina Feier. Ontology-based Choreography and Orchestration of WSMO Services [J]. <http://www.wsmo.org/TR/>.
- 34 Norton.B., and Pedrinaci.C. 3-Level Service Composition and Cashew: A Model for Orchestration and Choreography in Semantic Web Services, Workshop: 2nd International Workshop on Agents, Web Services and Ontologies Merging (AWeSOMe'06) .
- 35 James Scicluna, Axel Polleres, Dumitru Roman Ontology-based Choreography and Orchestration of WSMO Services. <http://www.wsmo.org/TR/>.
- 36 Domingue, J., Cabral, L., Galizia, S., and Motta, E. A Position statement on IRS III: A Comprehensive Approach to Creating and Using Semantic Web Services. Workshop: W3C Workshop on Frameworks for Semantics in Web Services, Innsbruck, Austria, 2005.
- 37 卢刘明. 基于语义的服务发现与组合关键技术研究. 中国科学院研究生院博士学位论文 2006,6.
- 38 Zhang Jianhong, Zhang Shensheng, Cao Jian, Mou Yujie. Improved HTN Planning Approach for Service Composition. Proceedings of the 2004 IEEE International Conference on Services Computing, 2004.
- 39 Xining Li, Wei Fan. An object-oriented And/Or graph inference engine Electrical and Computer Engineering. Canadian Conference on 14-17 Sept. 1993.
- 40 Walter Binder, Ion Constantinescu, Boi Faltings. Efficient Service Composition Using Zero-Suppressed Reduced Ordered Binary Decision Diagrams. Proceedings of the 2006. IEEE/WIC/ACM International Conference on Web Intelligence, 2006.
- 41 Evren Sirin, Bijan Parsia, Dan Wu, James Hendler, and Dana Nau. HTN planning for web service composition using SHOP2 [J].Journal of Web Semantics, 2004, 1 (4) :377-396.
- 42 Milanovic N, Malek M. Architectural support for automatic service composition[C]. In: Proc.of the IEEE Int'l Conf. on Services Computing (SCC 2005) .Orlando: IEEE, 2005, 133-140.

攻读硕士学位期间发表的学术论文

- 1 Jia Huitong, Wang Xin. Efficient Service Composition Architecture Using AND/OR Graph[C]. The proceeding of 1st International Conference on Intelligent Networks and Intelligent Systems, 2008. (EI 检索 20090111840168)
- 2 WANG Xin, JIA Hui-tong, Ma Yu-lian, Wang Yu-dong. Deviation Problem and its Solving [C]. The proceeding of 2008 International Symposium on Computer Science and Computational Technology[C], 2008. (EI 检索 20091211972277)
- 3 Wang Xin, Jia Huitong, Zang bohong. Semantic Web Composition of Deviation Problem. The proceeding of 2008 China-Ireland International Conference on Information and Communications Technologies[C], 2008. (EI Village,ISTP 检索)
- 4 贾慧彤, 王鑫. 基于知识的智能 Web 服务架构研究[J]. 现代图书情报技术,1003-3513, 总第 166 期,2008.
- 5 贾慧彤, 王鑫. 语义 Web 中 Trust 层与服务发现结合的评估模型[J]. 计算机应用研究.
- 6 郭艳庆, 贾慧彤. 语义 Web 服务桥体系结构—WSMO 桥深入探讨[C]. 中国智能系统工程学术大会.

致谢

衷心感谢我的导师王鑫老师的悉心指导与教诲。在三年的研究生学习期间，王老师给予了我耐心的指导和及时的帮助，非常感谢王老师对我在研究上的指导和督促，在思想上的教育，使我受益匪浅。读研期间所取得的研究成果，都是在王老师的悉心指导下完成的，它凝聚着王老师对我的精心培养和教导所付出的大量心血。

同时，感谢邱瑞华、周小兵、刘建丽、戴伟长、邓米克、蒋强荣等老师，他们严谨的研究作风、朴实的生活态度，都值得我去学习。

感谢实验室的师兄、师姐、师弟、师妹，以及和我一起实验室学习生活三年的兄弟们，向张伯宏、郭艳庆、尹珅、尚艳玲、刘健、赵娟等同学这三年来在学业和生活上对我的帮助表示感谢。感谢大家给我带来的辛苦而又快乐的三年研究生生活。

感谢所有曾经帮助关心爱护过我的人，谢谢！

最后感谢各位参加评审和参加答辩会的老师，感谢您们在百忙之中抽出时间来对我进行指导和帮助。