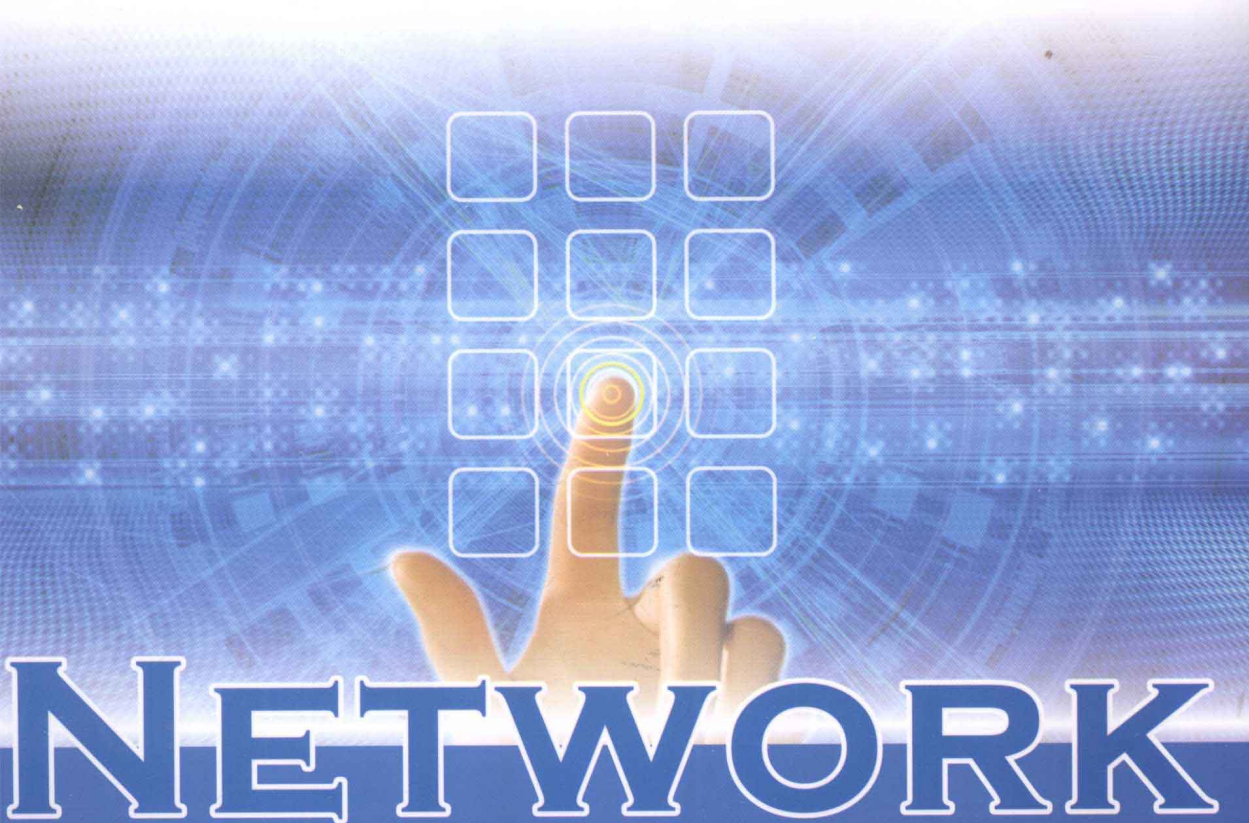


语义网

王莉◎著

社会网络计算 与Web资源共享



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

语义网、社会网络计算 与 Web 资源共享

王 莉 著

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书以当前 Web 主流技术和应用——语义网和社会网络软件为背景,以人们对资源共享的无止境追求为问题需求,参阅了许多经典学术文献和最新文献,在对相关国内外研究工作论述分析的基础上,介绍和分析了近几年相关研究领域的最新成果及作者的工作,内容新颖,可帮助 Web 研究和应用开发人员较快地了解相关领域知识。本书内容编排科学合理,结构清晰,语言通俗易懂,易于阅读和理解。

本书可作为理工科院校研究生或相关研究者的参考书,也可以作为企业研发人员了解 Web 资源共享技术的参考资料。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

语义网、社会网络计算与 Web 资源共享/王莉著. —北京:电子工业出版社, 2011.3

ISBN 978-7-121-12971-1

I. ①语… II. ①王… III. ①语义网络—研究 IV. ①TP18

中国版本图书馆 CIP 数据核字(2011)第 026334 号

策划编辑:秦绪军

责任编辑:徐蔷薇 特约编辑:王 纲

印 刷:三河市鑫金马印装有限公司

装 订:电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本:720×1000 1/16 印张:12.75 字数:236 千字

印 次:2011 年 3 月第 1 次印刷

定 价:39.90 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010) 88254888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010) 88258888。

前 言

在网络技术飞速发展、网络应用日益丰富的今天,网络资源共享成为一个永久的课题。不断出现的新技术,推动大量新型应用的产生,而新型应用又不断提出更高的技术需求,推动技术的飞速发展。现在的社会软件、对等架构、物联网、多网融合等应用,使得提供网络资源的门槛降低,海量的、动态变化的信息资源,文本、图像、音频、视频等多形式资源共存,真实和虚假掺杂的网络信息共现,实时信息资源的获取需求等,都对当前的 Web 资源共享提出了更高要求。

2001 年互联网之父 Tim Bernes-Lee 提出语义网的构想,为 Web 资源的智能联通和协作指出了一条路径。经过世界各国研究人员、公司研发人员近十年的努力,现在已经有许多的领域本体在不同行业域的信息集成和智能服务应用中发挥着重要作用,资源描述框架(RDF, Resource Description Framework)已经成为当前 Web 资源的通用表示框架,在一定程度上支持着更智能的 Web 服务应用。语义网时代的 Web 资源共享将会推动更智能、更个性化的 Web 服务应用。

即时通信软件、社会软件、C2C 电子商务等的产生,使得人类现实生活向网络延伸,而网络又以一种无处不在的面孔渗透到人类生活的方方面面,如电子邮件、即时通信、游戏、音视频欣赏、文献检索、协同科学研究、电子商务,以及近年来出现的物联网等,越来越多的网络用户原创信息内容充斥着网络,丰富 Web 信息资源的同时,也带来了更为复杂的网络资源管理问题。人类社会和网络社会从以前的映射关系转变为相互渗透和融合,Web 虚拟空间呈现出新的社会特征:一方面,由于不受地域和时间的限制,网络资源关系更多地是由人的内涵属性所决定的,更能揭示人类深层次心理特性之间的关系,显现出了以往社会组织中看不到的底层特征;另一方面,网络数据的可得性,使得研究人员发现了 Web 上的复杂网络特征。如何从宏观上发现网络资源结构特征,如何在微观上设计智能的资源组织管理机制,将成为优化资源组织的一个主要途径。

本书以当前发展迅速的语义网和基于 Internet 的社会网络为背景,对当前 Web 资源共享的若干关键问题和新技术进行了剖析和论述。本书主要分为 6 章,第 1 章阐述了当前 Web 资源现状、共享需求和语义网技术、社会网络及社会网络计算等。第 2 章介绍了一种实现 Web 资源共享的通信架构:知识通信。第 3 章介绍了基于语义网技术——本体的网络资源共享,从网络资源本体建模、描述

和语义相关性计算两个角度展开论述。第4章对知识资源的虚拟组织结构研究进行了阐述,综合分析了复杂网络的重叠社区结构探测、动态社区发现、自主语义社区自组织形成等研究工作,介绍了一种基于共享邻居关系的社区结构发现算法、多图转换的动态社区发现算法,同时在微观层介绍了一种动态对等网络分布环境中基于信任和被信任关系的自组织社区形成机制,以适应存在大量资源的复杂动态网络环境中资源的自组织管理。第5章对支持资源共享的知识通信协议的研究工作进行了介绍,详细阐述了一种能感知内容和上下文的适应性通信协议。第6章介绍了一种知识通信典型应用系统——高等教育资源语义共享平台。

本书是在国家自然科学基金项目“知晓内容和环境的本体知识路由研究”(60472093)、山西省青年科学基金项目“内容和上下文感知的动态适应性协议研究”(2006021015)、山西省回国留学基金项目“动态多模复杂社会网络中社区发现及演化的关键技术研究”(2010-31)三个项目资助下的一个综合成果,也包含了作者近年来的主要研究成果。

在此,要特别感谢太原科技大学的张继福教授,他认真阅读了本书全稿,提出了许多极有价值的修改意见。同时,本书的出版得到了山西省人社厅留学回国择优项目和电子工业出版社的大力支持,在此一并致以真诚的感谢。

由于作者水平有限,不足之处在所难免,欢迎读者批评指正。

目 录

第 1 章 引论	1
1.1 网络资源	1
1.2 Web 资源共享	2
1.2.1 Web 应用与资源共享	2
1.2.2 Web 资源表示和互联互通	3
1.2.3 Web 资源组织和管理	3
1.2.4 网络架构与资源传递	6
1.3 语义网	7
1.4 社会网络和社会网络计算	8
1.4.1 信息资源网络与社会网络的关系	8
1.4.2 社会网络	8
1.4.3 复杂网络	10
1.4.4 社会网络计算	15
1.5 技术发展带来的新需求和新契机	16
第 2 章 知识通信：一种实现 Web 资源共享的基础架构	18
2.1 数据、信息和知识	18
2.2 知识通信	21
2.3 基于动态虚拟语义社区的知识通信	23
第 3 章 语义网技术支持的 Web 资源语义共享	26
3.1 异构资源语义共享	27
3.1.1 Web 网络资源模型	27
3.1.2 本体	30
3.2 面向领域元数据的多维语义共享资源本体模型	32
3.2.1 问题背景	32
3.2.2 本体驱动的网络资源多维语义共享模型 RSM	34

3.2.3	网络资源信息互联的实现——资源表示规范的本体建设	35
3.2.4	网络资源知识互联的建设——领域内容本体建设	36
3.2.5	网络资源知识协作——协作过程本体建设	37
3.3	基于本体的异构语义计算方法	41
3.4	一种本体关系驱动的 OWL 概念语义相关计算	46
3.4.1	OWL 类间关系的形式化模型	47
3.4.2	类间关系公理	49
3.4.3	问题模型及相关引理	52
3.4.4	基于本体结构和语义关系的语义相关度计算	53
3.4.5	实验及分析	56
3.5	本章小结	59
第 4 章	基于社会网络计算的网路资源组织	60
4.1	社会网络、知识通信与社区	60
4.2	社区结构	61
4.2.1	社区	61
4.2.2	社区发现算法	62
4.2.3	社区模块度	67
4.2.4	重叠社区和动态演化社区	68
4.3	重叠社区发现	69
4.3.1	重叠社区相关研究	69
4.3.2	基于共享邻居关系分析的分层重叠社区发现算法	71
4.3.3	NHOC 算法实验及评估	77
4.3.4	一种支持粒度优化的重叠社区质量评价	79
4.3.5	重叠社区质量评价模型实验及分析	84
4.4	动态演化社区发现	92
4.4.1	复杂社会网络中动态社区研究现状及分析	92
4.4.2	问题模型与经验假设	95
4.4.3	邻居域信息驱动的多图变换的动态社区发现算法	97
4.4.4	基本符号表示和相关定义	99
4.4.5	面向 0-1 网络的动态社区发现算法	100
4.4.6	面向带权网络的动态社区发现算法	101
4.4.7	动态社区发现算法的相关计算	102

4.4.8	实验及评估	104
4.5	对等网络与社区	112
4.5.1	社会网络和对等网络	112
4.5.2	基于社区虚拟组织的对等网络资源定位	113
4.5.3	基于社区虚拟组织的对等网络资源质量管理	116
4.5.4	P2P 社区研究分析	117
4.6	基于信任评价和专家驱动的非结构 P2P 自主语义社区模型	117
4.6.1	P2P 中的社会关系观察与分析	117
4.6.2	模型的定义和表示	119
4.6.3	自主节点本地信任评价	123
4.6.4	自主语义社区形成	125
4.6.5	全局信任融合算法	128
4.6.6	本地路由表的动态调整及路由机制	130
4.6.7	分析及实验	133
4.6.8	讨论	137
4.7	本章小结	139
第 5 章	智能通信协议	140
5.1	Agent 知识通信协议	141
5.1.1	Agent 通信目的和需要解决的基本问题	141
5.1.2	Agent 知识通信语言 KQML 简介	142
5.2	语义 Web Services 的通信协议	149
5.3	基于本体的知识通信协议	154
5.4	内容和上下文感知的知识驱动的通信协议	155
5.4.1	研究动机	156
5.4.2	内容和上下文感知的通信共享知识库——KC ² A ² P 协议本体	157
5.4.3	随需随境的协议编码机制	161
5.4.4	内容和上下文感知实验	163
5.4.5	知识通信协议的适应性软件架构	165
5.5	本章小结	166
第 6 章	高等教育资源语义共享平台	167
6.1	高等教育资源共享平台与知识通信	167

6.2 高等教育资源语义共享的实现	168
6.2.1 网络资源语义模型 RSM 在教育资源共享平台中的应用	168
6.2.2 资源语义映射和集成	173
6.2.3 系统实现与评估	175
6.3 语义开发工具选择	177
6.3.1 本体编辑工具——Protégé 简介	177
6.3.2 本体应用 API-Jena 简介	178
6.4 本章小结	179
参考文献	180

第 1 章 引 论

1.1 网络资源

从来没有哪种传媒工具能像互联网一样，以如此快的速度渗透到人类生活中，并以一种不可抗拒的力量不断改变着人们的思维、习惯和生活。据中国互联网络信息中心（CNNIC）发布的《第 26 次中国互联网络发展状况统计报告》显示，截至 2010 年 6 月底，中国网民规模达 4.2 亿人，互联网普及率持续上升至 31.8%。技术的发展和创造性应用的不断产生，以及网络资源的极大丰富，推动了网络社会和人类社会的极大融合，互联网带给人们的娱乐功能、商务功能、通信功能，以及现代物联网将带来的全新智能生活空间功能，正在不断地改变着人们生活的方方面面。

当前计算机网络技术和应用的现状是：

（1）出现了许多新型计算模式，如网格、对等计算、Agent 技术、普适计算、云计算等。现代网络应用开始注重开放环境中的个体间的协同计算。

（2）涌现了许多新的网络应用模式，如维基百科、博客、日志、电子新闻、网络虚拟社区、网络论坛、即时通信和电子商务等。

这些新型应用架构使互联网资源的提供和传播从以服务提供商为主逐渐转变为以网络用户为核心，资源的提供和使用门槛降低，为计算机网络资源共享带来了蓬勃发展生机，也带来了新的挑战：

（1）资源来源广泛，大量的分布异构资源不断加入、更新与扩展，质量良莠不齐，资源内容观点有对有错，信息有真实有欺诈，服务质量有好有坏。

（2）资源数量飞速增长，据维基百科介绍，2008 年 7 月，Google 可以搜索到的不重复网页数已达到 1 兆万亿页，每天还以新增几十亿页索引的速度增长，而 Web 静态网页的实际数量可能是 Google 可索引网页的近 10 倍。海量数据和信息在丰富网络资源的同时，也增加了快速找到真正所需资源的难度。

（3）资源形式多样化，有软件、文本、数据库、音频、图形图像、表格和动画等各种形式。

(4) 资源间关联关系也在不断动态变化。

(5) 计算机网络资源显现出一定的结构特征和社会属性。主要体现在两个方面：一方面，Web 网站的设计者们在建立其链接关系时，总是会选择热点网站或链接度高的网站，人类的“富者越富，穷者越穷”的偏好选择（马太效应）在网络分布中得到了体现；同时，有研究者发现，人类社会的小世界现象也同样出现在网络分布中。另一方面，许多社会软件的出现和广泛应用，使得计算机网络，尤其是 Web 网络，已经从最初的人类社会的辅助设施逐渐形成一个与现实世界相对应的 Web 世界（Web Wide World），其作为现实世界的映像，与现实世界紧密地融合在一起，成为推动并主导人类社会发展的越来越重要的因素。网络已经从仅仅呈现静态信息资源的载体，发展演变为多模式、多资源动态链接、交互、协作的平台。

1.2 Web资源共享

作为互联网的重要应用之一——万维网 Web 自产生之日起就显现出蓬勃的生命力。从最初页面互联、信息互联的 Web1.0 到现在的知识互联 Web，从人人互联的以社会软件为主体的 Web2.0 到将来的智能互联 Web3.0，无论是哪个发展阶段，推动 Web 不断向前发展的主要动力就是资源共享、集成和协同的更高需求。

1.2.1 Web 应用与资源共享

当前，Web 上的应用主要有通信功能、娱乐功能、商务功能、科学研究功能等，其中，媒体文件共享、在线电子商务、搜索引擎、各种在线社区等成为主要功能增长点。

在媒体文件共享中，为了快速、准确地给用户提供其所需要的高质量资源，必须对不同类型的媒体文件进行有效的内容管理和质量管理。在搜索引擎中，需要对抓取到的海量 Web 资源进行聚集、分类、组织和更新等。在现代全球大力推进的扩展型网络应用——物联网中，为了实现物理世界、人类社会、信息社会的融合，需要多源软件的协同、多源数据的有效集成。由此，资源共享协同问题成为各类网络应用中急需解决的关键问题。

总之,无论哪种网络应用,智能服务是它们共同的最终目标,智能服务离不开资源间的智能协作,而智能协作必须要有资源共享作为其基本支撑。可以看出,无障碍的资源互通、有机的资源组织、高效的资源传递将是实现资源共享的一种有机架构。

1.2.2 Web 资源表示和互联互通

网络资源的表现形式多样、开放和动态变化引发了资源共享中的第一个关键问题——资源异构性问题,它是网络资源互通、集成和共享协同的主要瓶颈。当前网络技术的发展已实现了资源间的硬连通,阻碍深层次资源连通的问题主要是资源语义异构问题。因此,需要在适应当前网络资源多样性特征和拓扑的情况下,建立资源语义互通方法,以支持资源的集成与共享。通过对当前相关研究的分析,我们发现学者们一般从以下三个方面进行研究:研究网络资源的表示和描述方法,以表现出资源的全面语义特征;设计算法,让具有不同描述形式的(不同表示语言的)资源间能相互识别及资源集成;设计机制,在不改变现有网络结构的情况下,使异构系统间能够共享资源。只有资源语义互通,异构资源才可能实现真正共享,从而提升网络服务质量。其中,资源表示的结构化和知识化处理、语义相似计算等成为当前解决网络资源异构性的主要途径之一。

1.2.3 Web 资源组织和管理

资源语义互联互通使得分布异构多样的网络资源共享协作、集成成为可能,但这仅是提升网络效能的最基本的第一步工作。网络资源的海量性、资源质量的多样性,需要有效的资源组织管理机制来提升网络服务质量。以当前网络服务的主要应用之一——搜索引擎为例,先进的计算方法和技术使得网络可到达多种资源并可抓取回资源信息,但人们仍然需要在返回的大量信息中逐条进行判断。

人类在认知事物时,喜欢按照分类体系进行;而自然界和社会事物在分布时,总会呈现出一定的“物以类聚,人以群分”的聚集性。所以,从表面无序的资源分布中发现有序特征并加以利用,建立网络资源的再组织管理体系或虚拟组织,可有效提高网络资源定位和路由效率。另外,网络的动态开放性使得无法界定固定的资源管理边界,网络向人类生活的渗透所引发资源的社会性使得资源具有局部自主性,这些使得运用集中控制管理的方法对资源进行组织已远不能满足高效

充分使用资源的要求。因此,现在的网络资源还需要分散、灵活的组织管理,以及灵活的资源协作体系。现代计算科学和数据挖掘技术的发展,以及网络数据的可收集性等,使得研究者们可以通过观察和统计分析,得到网络资源的分布特征,然后基于其宏观特征,设计新型的资源管理模式,由宏观引导微观,微观反过来又影响宏观,使得网络资源之间可以更好地协作。

具体来讲,当前 Web 资源管理可以分为以下四种模式。

1. 信息资源的直接管理

信息资源的直接管理是一种直接面向网络信息资源单元的组织管理方式,它主要表现为文件、自由文本、超文本、网页或网站等组织管理方式,是最简单,也是粒度最小的 Web 资源管理方式。

2. 信息资源的再组织管理

信息资源的再组织管理是对资源直接管理模式所组织信息的再组织。根据该再组织体系对用户可见还是不可见,信息资源的再组织管理可以分为两大类:一类是用户可见的组织管理模式,典型代表就是目录指南方式。目录指南方式也称为主题树方式,它将信息资源按照事先确定的概念体系分门别类地逐层加以组织,然后显式地提交给用户,用户先通过浏览的方式层层遍历,直到找到所需要信息的线索,再通过信息线索链接到相应的网络信息资源。它的缺点是必须事先建立一套完整的主题目录分类体系,而且用户必须对分类体系有一定的认识,同时,为保证主题树的清晰性和资源的可用性,主题目录体系结构又不宜过于复杂,每一类目下的信息索引条目也不宜过多,这就与体系的完整性产生了冲突。

另一类组织管理模式是用户不可见的,搜索引擎是这类模式的一个典型代表。搜索引擎是一种 Internet 网络服务,它的主要任务是在 Internet 中主动搜索其他 Web 站点中的信息并对其自动索引,其索引内容存储在可供查询的大型数据库中。当用户利用关键字查询时,该网站会告诉用户包含该关键字信息的所有网址,并提供通向该网站的链接。它的资源组织方式对用户来说是不可见的,其优点是能够高速处理大量结构化和非结构化的信息,对于信息处理也更加规范化,特别是在信息大数据量的环境下,其优点更为突出,但它对用户提出了一定的要求,要求用户必须掌握一定的检索技巧,包括关键词及其组配的选择。

3. 信息资源的内容重组管理

信息资源的内容重组管理是在对源信息所含知识内容进行分析解构的基础上,运用一定的科学方法将源信息进行重新组合,从而得到新的信息产品,实现信息增值的过程。例如,返回给用户查询的结果不再像传统搜索引擎一样,提供的是不同链接的列表,而是一个结果页面,其中包含了不同来源网站或网页中满足查询需求部分的内容组合等,这也是下一代智能搜索引擎的服务目标之一。

4. 自主协作的资源组织模式

互联网是无中心控制的分布式网络,在表面无组织的网络运行过程中,经常会隐含、甚至会涌现出一些有序的组织结构,这些有序的组织结构对于保持网络稳定、鲁棒性,以及发现网络中的重要信息都具有非常重要的价值。所以,在网络资源管理中,除了上述三种管理模式之外,许多研究者们开始从网络资源组织结构发现的角度进行研究,尤其是许多新兴计算结构和应用模式的出现,使得网络资源更加丰富和多元化,网络资源间的关系也更丰富和复杂,因此,通过对网络资源间可见的关系进行分析,发现资源间隐含关系,引导资源组织和有序构造,将有效提高资源的使用。这方面的研究随着网络数据的可得性、计算技术的进步等成为当前研究的热点。

这里以网络上新兴资源 Blog 和新型计算模式对等结构为例进行介绍。

Blog 是继 E-mail、BBS、ICQ 之后出现的第四种网络交流工具,是一种借助于网络来表达个人思想的出版方式,它其实就是一个网页,通常是由链接和简短且经常更新的帖子所构成,这些发布的文章都按照年份和日期倒序排列。Blog 资源提供主体是社会的人,资源在网络上所呈现的关系隐含着人的知识、认知及社会结构信息。由于 Blog 资源的提供门槛低,数量飞速增长,还有 Blog 所隐含的人类社会信息,使其成为网络上的一种重要资源,社会学、计算机界等不同领域的研究者们开展了 Blog 结构关系发现的研究,其主要研究方法就是通过链接关系分析,实现对 Blog 资源的组织管理和基于 Blog 的社会关系发现。

对等结构模式(P2P, Peer to Peer)是一种架构在互联网上的新兴计算模式,其节点间关系是平等的,即没有客户端和服务端之分,资源提供者同时也是资源使用者,所以, P2P 网络中资源数量大、内容丰富,许多文件共享、视频共享、即时通信等都是基于这种模式。但是由于 P2P 节点具有自私性和自主性,资源质量混杂、变化快、数据量大等特点,使得资源管理和使用成为对等结构在更多、

更广领域发挥效能的瓶颈。同时,它的节点角色的平等性使得网络系统不可能从全局角度进行资源组织管理的整体控制,需要设定局部策略,使节点间协作通信,从而构成一定的有序结构。这种自主构造的结构有时是显式显现的,有时呈隐式,对这种结构进行发现并进行引导,可以有效提高网络资源的服务效能。

可以看出,通过制定标准规范,建立一定的搜索、筛选、分析、排序、存储等手段使分散、无序的信息形成一个有序的结构;发现资源隐含关系,设计资源协作策略,促进资源自主形成有序结构,这两种组织管理方式对当前网络资源管理和使用都具有重要意义。第一种研究是从网络资源表现形式对资源进行的组织管理;第二种研究将挖掘资源背后的特殊属性,尤其是其社会属性、认知属性和意识属性等,通过对其可见特征分布的分析,建立其分布模型,并在微观层面建立局部节点策略,促进协作和有序结构的生成,以利于资源的高效利用。两类研究中,前者着重于主动的全局设计和规范,后者着重于适应性地发现和局部设计,两者相辅相成,将大大提升资源管理和使用的效能。

这也成为本书内容结构组织的一个重要依据。在本书所介绍的知识通信架构中,通过探讨资源模型语义表示规范及语义计算等问题,建立显式的信息门户模式(我们称之为“资源共享平台”),实现了标准规范支持的 Web 资源组织管理;同时,通过发现网络信息资源的宏观拓扑规律,构造微观自组织协作构造方式,宏观指导微观,微观适应宏观并促进宏观发展,实现了灵活可扩展的 Web 资源自主管理和使用功能。

由于对等结构的广泛应用和对等结构有待挖掘的巨大潜能,本书中的自主协作的网络资源组织方式研究主要以对等结构为背景展开。

1.2.4 网络架构与资源传递

实现网络资源的互联互通,并在此基础上建立或构造良结构的资源管理,就好像各种类型的交通子系统(公路、铁路、水路、航空等)具有统一地图描述,并且根据地域等信息进行了有效组织和联通,但是要想在这样的交通网络中实现快速、准确的运行,还需要优化的路径规划。在路径规划中,必要信息的获取和传递是路径规划质量保证的前提。网络资源的使用也是如此,资源互通和资源组织管理,构成了可高效利用资源的基础设施,但还需要高效的资源信息传递,才能最终实现高质量服务。通信协议是通信各方事先约定的规则,用来实现网

络连接。要实现网络资源共享,支持不同资源间互联互通的协议是不可缺少的部分。在现代网络技术中,TCP/IP 协议使得不同终端的计算机设备得以相联相通,但是新兴网络应用对其提出了许多智能适应性功能需求。例如,无线网络、移动网络、Ad hoc 网络、汽车平台通信网络等动态网络的快速发展,需要智能协议体系等支持其动态复杂环境中信息的准确、快速传递。由此,许多对内容感知、环境感知的路由协议框架成为当前研究的热点。

语义互通、资源组织和智能协议是当前应用层网络资源共享研究领域的三个重要问题,其关注点各不相同。但这三个问题又是环环相扣,具有一定的内在联系的,前者可以作为后者的基础设施,后者是基于前者的进一步工作。

已有相关研究工作一般都是面向网络层,而现代新型网络技术及计算机技术的发展,为在更高层次实现新的功能、提升服务质量提供了新的途径。

1.3 语义网

2001 年万维网的创始人 Tim Bernes-Lee 等人提出语义网,其目标是建立数据 Web,以提供跨越应用、企业、社区等界限的数据共享和重用的公共框架^[1]。他们所提出的语义网分层标准栈中的 XML (Extensible Markup Language)、RDF (Resource Description Framework)、RDFS (RDF Schema)、OWL (Ontology Web Language)等提供了从仅描述数据结构到描述数据模型及语义的不同层次的网络资源表示方法。语义网具有以下本质特征^[2]:

- (1) 通过万维网能访问到原来仅供本地使用的应用信息。
- (2) 以基于文档结构的形式化描述促进数据集成和组合。
- (3) 以机器理解的形式描述信息,使智能 Agents 能理解信息并根据信息进行操作。
- (4) 建立了数据和真实世界对象间的关系。
- (5) 可以利用作用在数据上的逻辑规则推出新的知识。

经过近些年的发展,有学者对语义网分层架构提出质疑,但是,XML、RDF、本体等作为资源表示的方法已经被广泛应用^[3],它们定义了机器可理解的类及属性等,并且用唯一的通用资源标志符进行资源标注,这些为在更高层次根据内容进行异构系统交互、资源互联和共享、提升网络资源发现、传递质量提供了新的途径^[4]。

1.4 社会网络和社会网络计算

1.4.1 信息资源网络与社会网络的关系

网络技术的迅猛发展和网络应用向人类生活的广泛渗透,以及新兴网络模式的产生,迅速增长起来的维基百科、Blog、网络社区、E-mail、BBS、即时通信等使得资源形式更为多样,资源间关系更为复杂,资源网络拓扑动态性更强。但同时也可以看出,网络所显式表现出的资源间的关联关系,其实都是资源提供者的知识结构,以及社会人对网络资源认知状况的体现。例如,在 Blog 的链接中, Blog 制作者倾向于把自己的 Blog 链接到朋友、权威站点或具有相同主题的 Blog 上;而且,虽然资源自身不具有思维意识及意愿,但资源背后的人具有这一切自主性,所以,网络上也会存在虚假信息资源、低质量信息资源甚至欺诈信息等。因此,在对资源组织管理时,表面上我们面对的是资源实体,实际上真正的对象是社会人。资源间的关系,表现的是社会人的知识关系及认知状态,所以,对资源组织管理机制的研究,从深层次讲,就是社会人在以计算机为媒介的网络世界中的认知、知识及行为的组织管理。对网络资源结构的发现研究可以转化为对基于计算机网络的社会网络的研究。

1.4.2 社会网络

根据维基百科的解释,社会网络(SN, Social Networking)来自于社会学理论,是指个人之间的关系网络。社会网络的主体是人,人之间的各种联系构成了关系,这种关系既包括静态的、由某种社会结构所决定的关系,如家族血缘关系、行政等级关系、企业组织机构关系和由地域、爱好、利益等相似性所引发的聚团关系等,还包括动态的、基于上下文不同而随时变化的、由交互和活动等动作所引发的动态关联关系等。

社会网络理论中有两个非常重要的基础理论*:六度分隔理论(Six Degrees of

*引自网络营销教学网站 (<http://www.wm23.com>)。