

公共信息整合策略研究综述^{*}

张道顺 白庆华

(同济大学经济与管理学院 上海 200092)

摘要 整合已经成为各组织管理中的主要矛盾。面对无法逆转的社会主体对政府公共服务一体化需求,政府部门能否在公共领域提供畅通的服务,就成为信息化建设过程中公民关注的焦点,公共信息整合是电子政务(e-Gov)建设中亟待解决的现实问题,本文从政府门户、信息分类整合等方面对公共信息整合进行了综述。

关键词 信息整合,电子政务,政府门户,元数据,互操作性

An Overview on Public Information Integration Strategy

ZHANG Dao-Shun BAI Qing-Hua

(The School of Economics and Management, Tongji University, Shanghai 200092)

Abstract Integration has become the main contradictions for any organization in management. Facing the holistic demands from the public which are not changed, it is the focus concerned for government whether it can provide seamless information services or not for the public during the construction of e-Gov, so it is the urgent task for government to put its energy to integrate all kinds of public information. In this paper we give an overview on the public information integration strategy from some aspects, such as e-gov portal, public information categories and so on.

Keywords Information integration, E-Gov, Government portal, Metadata, Interoperability

0 背景

在全球化背景下,资源在局部范围应用发展到在全球范围内进行合理配置,人们对物质、能量和信息资源认识的视角已从组织内放眼到组织外,什么是资源最佳配置?如何协调促进资源在不同组织组织内的合理流动,创造更多的物质财富?无论是客户关系 CRM、供应链管理(SCM)、电子商务(EC)、还是人力资源管理、敏捷制造、企业应用集成 EAI 等以及由此而衍生的其他的管理理念无不是从整体、全局的角度来理解、认识资源、组织、生产、流通、消费、发展之间的关系,也就是资源整合的思想,各种管理思想莫不是将信息作为精髓、指导,研究信息与其他要素之间的关系及其合理流动,借助信息合理流动实现资源最佳配置。“整合”这一概念被应用于各个领域经营管理之中,成为互互联网中出现频率最高的词汇之一,并成为组织管理中的主要矛盾。

公共领域信息化主要表现为电子政务(e-Gov)建设。从信息抽象的角度来认识政府治理,其过程实质就是政府广泛搜集经济、自然和社会复杂系统的各类信息,再进行加工整理后向社会发布有关指令性、指导性、调控性、解释性、服务性信息及处理过程,同时快速而有效地收集来自各方面反馈信息,进而对现有政策、法规和措施进行修正和优化。e-Gov 已经成为公共管理通向信息社会的标界和路由^[1],政府是信息时代信息及信息技术最大的消费者,在我国有 80%的信息是由政府部门掌握的,面对政府信息管理错综复杂的环境,面对无法逆转的社会主体对政府服务一体化需求,政府能否在公共领域提供畅通的服务,就成为信息化建设过程中公民关注的焦点,成为衡量一个国家和地区的信息化水平的重要指标,也就是摆在政府部门面前如何整合为公众所有、而由公众信任

的政府部门创建、编辑、编译和维护的公共信息^[2],为公众提供满意的信息服务。

1 公共信息整合的必要性

知识经济时代的管理实际上是运用信息和信息技术创造概念和行为对象的过程,追求的是独特的核心竞争力。随着国家信息化、区域信息化、城市信息化、社区信息化、企业信息化的不断深入,信息技术已经不仅仅是在局部领域的应用,随着认识的不断深入,信息渗透并被应用到社会的各个层面,信息化已经成为一个系列、一种群体和一个体系行为,成为一个国家追求的目标,公共领域的信息及信息管理对一个国家、地区的发展及其营商环境起着至关重要的作用,其水平高低已经成为衡量一个国家现代化水平和综合国力的重要标志。

世界范围内各组织都面临着信息集成整合的考验^[3],一是整合遗留系统(Legacy System)数据的需要;二是需要集成第三方或战略伙伴系统数据到本系统中,三是需要对分散的数据源的数据进行分析。在美国企业咨询委员会 EAC 向 OMB 递交的报告“互操作战略:概念,挑战和建议”^[4]中指出:“政府是信息技术最大的用户,估计每年有 500 亿美元用于投资 IT,其中有相当大的一部分用于遗留系统的维护开发支持,大部分投入不是为了系统的协同工作,我们的系统和数据库被构建成仓筒式的结构,政府所面临的最大挑战之一是使这些系统互操作和共享信息”。在英国政府的互操作框架 E-GIF (E-government Interoperability Framework)^[5]中指出:联合的政府需要联合的信息系统(Joined-up government needs joined-up information system),在英美等信息化水较为发达的政府的影响下,一些国家包括亚洲的新西兰等国也开始制定本国电子政府的互操作框架^[6]。由此可见,信息发达国

^{*}上海市科委自然科学基金项目“城市信息化中信息整合研究”部分成果。张道顺 博士研究生,研究兴趣:城市信息化、电子政务。白庆华 教授,博士生导师。

家政府已经将信息整合作为一种战略来抓,从宏观上来规范、引导正在进行又一轮信息化建设,并将此作为一种自觉的行为。

技术的发展以及私营组织的影响,社会主体对公共信息服务提出了深层次的需要。公众对政府部门创建的信息的所有权已经成为公众最基本的权利^[2],渴望公共服务部门提供“one-stop, non-stop”式透明服务,屏蔽政府部门间的组织界限,然而从过去旧的管理体制、机制、技术、标准中建立起来的各类信息系统、信息资源越来越难以满足人们的需要。不同网络环境、多重分散的操作系统、不同的数据库管理系统、不同的数据格式、不同元数据方案(Metadata Schemes)、不同的术语(Vocabularies)和本体(Ontologies)、不同的领域(Disciplines)和不同语言、不同的字符集、不同的检索系统、不同信息发布方式、系统重复、条块分割以及与信息发展相左的管理者理念意识^[7],政府部门公共信息一方面在快速增长,另一方面由于缺乏相应的管理机制,政府发布在 Web 上公共信息也在快速消失,上述因素阻碍了信息共享及其功能进一步发挥,成为信息化建设的制约,从而使得公共信息整合成为信息化建设中的战略问题之一,成为电子政务建设中亟待解决的现实问题。

从信息系统所涵盖的要素来看,可以从不同的视角来审视公共信息整合。对用户而言,由于信息知识、技能、收入层次等差别,希望政府部门能够屏蔽政府内部组织结构的复杂性和信息的广泛性,为公众提供集成统一全局领域的信息视图和无技能的透明服务;从系统这一视角,要解决不同系统之间不兼容的矛盾,实现系统之间的交换、互操作性和共享;从数据和信息角度来认识信息整合,不仅因为数据和信息分布于不同的系统,而且因为数据和信息具有不同的结构、不同的格式、不同的语义,所以信息整合就是要解决数据信息一致性和信息语义问题;从信息流动的动态地来考虑信息整合,就是要实现信息流(服务流)准确、快速流动和相互之间的协调;从信息维护管理角度,就是要解决信息海量性、动态性、分布性等与信息管理的集中性和便捷性的矛盾。总之,公共信息整合目的是要实现分布式环境中在正确的时间、以正确的方式,将正确的信息传递给正确的用户。本文将从政府门户、公共信息互操作等五个方面进行综述。

2 电子政府及门户建设(平台和业务的集成整合)

电子政府是当前公共信息领域信息整合的最佳机遇,由于新公共管理理论(NPM, New Public Management)对电子政务建设的深远影响,公共服务领域中的一站式或一门式服务成为电子政府建设的目标和公民最强烈的呼声,通过集中式门户网站将分散于政府各组织机构的信息服务的入口整合成一个统一的、为公众提供多访问渠道的、直达各政府机构提供的集成一站式信息服务,屏蔽政府内部操作的复杂性,政府门户是公共信息服务的集成平台,其建设已经成为政府整合公共信息服务的重要策略。为推进欧洲电子政务的建设,EC(European Commission)建立并提供资金支持数项 e-Government 计划,其中之一是 Egov(An Integrated Platform for Realizing Online One-stop Government)^[1],这项始于 2001 年 6 月的为期两年的研究计划,主要目标是规范、开发、利用和评价实现联机一站式电子政府的整合平台。这一集成的平台允许公共部门为公民、企业伙伴和政府公务员提供基于生命事件和业务状况的信息服务,它的技术目标是要规范和开发:(1)下一代一站式政府门户和支持架构(包括 WAP-访问、个性化、定制服务、多语言支持、推送服务和识字签名);(2)服务

库(SR, Service Repository)和服务创建环境(SCE, Service Creation Environment);(3)政府标记语言(Government Markup Language),由于 XML 语言灵活性以及数据表示和应用相分离的特征使其成为构建其他语言的元语言,通过在 XML 基础上构建的 GML 实现政府门户与政府各信息和服务库系统的连接和数据互换和信息共享。

政府门户绝不是简单的服务入口的整合统一,它还涉及到政府元数据标准、政府系统互操作战略的制定,涉及到中央政府与各地方政府之间服务协调、应用服务的注册登记、相互责任的界定、政府网关等,服务是以组件的方式通过网络提供给任何位置的任何人,也就是分布式的服务能够与其他基于服务的应用进行交互,政府门户的注册网关(Registry Gateway)成为最重要的基础设施^[4,5],Registry 为跨部门协调命名空间、专用术语以及数据标准提供手段,这种 Registry 是一个逻辑注册中心,或称之为 Registry System,Registry 将会使电子政府受益匪浅,它的优势在于:允许流程的发现、促进复用减少劳动、有效版本控制、跟踪多版本业务对象、促进对注册对象的统一理解、合作、服务的统一调整、安全控制等。

Karen Layne 和 Jungwoo Lee 以欧美 e-Government 实施的背景提出了电子政务建设的四个阶段^[8]:①政府信息门户;②政府在线事务处理;③政府业务垂直整合;④政府业务水平整合。垂直整合是把不同级别的政府在相同功能上的整合,水平整合是把不同功能和服务在同一级别的政府上的整合。Karen Layne 认为在阶段③和④会形成一些国家级的集中的数据库(物理或逻辑集中)。

美国联邦政府的企业架构(Enterprise Architecture)是通过将业务以可分割的组件进行适当的划分和松散耦合,这样结果对应用的其他部分将会产生最小的影响^[5]。使政府信息整合可以在三个层面进行:(1)数据层;(2)活动流程层;(3)用户接口层。这种架构具有:用业务术语和目标快速规范受控业务操作的变革能力、提供实施服务的能力,消除开发者为解决相同问题重复设计和编写软件组件的重复劳动、让用户集中于目标的实现,而不必去理解所使用的各类工具。

3 公共信息互操作整合策略

互操作是网络环境中信息发现和检索的根本挑战,在文献中通常被当作一个完整的整体来对待^[9],实际上互操作是多层面的,可以分解成不同的类型和层次水平。William E. Moen 认为互操作要从系统和用户两个方面来认识,从系统方面是指系统或部件之间无需特别的要求可以进行信息互换和使用互换信息;从用户的角度,互操作是用户以有意义的方式搜索和检索信息并对所得结果充满信心的能力。要实现互操作,组织应该积极参与正在进行的过程,确保组织的系统、过程和文化以这样的方式在管理,那就是无论是在组织内或组织间最大化信息互换和重用机遇^[10]。由此,除了兼容的软、硬件外,真正实现互操作还有许多工作要做,有效的互操作需要组织工作方式的变革,特别是组织的信息态度。互操作从 IT 的角度是维护参与者(互换信息)的自治性,允许各自系统维护自己的词汇术语和计算机环境。Paul Miller 将互操作分成以下几类:技术互操作、语义互操作、政治的/人员互操作、团体内的互操作、合法的互操作、国际范围互操作^[10]。William Moen 认为互操作与机构团体之间的共同性的程度密切相关,他还同时给出了机构共同特征和异质性的一些层次划分,机构可以分成三类^[11]:Focal、Extended 和 Extra,组织共同特征越少,互操作的难度就越大。

互操作已经从纯技术研究上升到信息化战略高度研究,

已经成为一个体系,英美等信息化发达国家和亚洲的新西兰都制定了本国的政府互操作战略,英国政府的互操作框架已经推出了第五个版本,E-GIF 陈述为实现跨部门互操作的政府的政策和标准。

新西兰政府 2002 年 2 月发布了电子政府互操作框架 (VERSION 0.9 和 RELEASE 8),在 2001 年 12 月发布的第二版新西兰电子政府战略中对互操作定义为:“通过共同标准共享信息、集成信息和业务的政府能力”,它由五个部分组成,即业务流程接口、服务传送、访问、信息共享互换(数据集成)、相互连接。实现互操作:系统之间互换结构化数据,互换元数据,互换和继承业务流程,元数据的提供和搜集,互换各类型文档。

E-GIF 为电子政府提供了公共基础环境,是电子政府的核心。可以让每个机构利用已有的方法和标准都加入到电子政府建设中,是电子政府成功的关键,电子政府的互操作框架的使用,能够促进跨组织边界的信息和服务的整合,为公众提供了对政府信息简便的访问方式。

Lubash 从 Specification、Workflow、Contract、Presentation、Relationship、Directory Service 等方面提出了互操作金字塔型信息架构^[4],给出了实现互操作较为全面的体系架构。

Z39.50 是分布式网络信息检索协议,该标准源于 1984 年,经过多次修改于 1998 年成为国际标准 ISO2395^[12]。Z39.50 的互操作主要集中在三个层面:协议语法层、协议服务层和语义层,其主要功能表现为:一个用户接口可以访问多重著录和数据库、一次会话中可以搜索多个著录系统和数据库、从多个服务器搜索的结果可以一标准的格式返回并可以进行数据的组合、结果可以被 Client 端重用、可以对结果和数据项实施排序等工作。

作为成熟的标准,Z39.50 被广泛应用于商用数据库、图书馆、博物馆和政府信息系统建设,尤其是被用于实施互操作,在政府领域,Z39.50 已经被用于 GILS 体系和 Clearing House。

文[4]中指出,为实现政府一体化的企业架构的解决方案,必须要解决信息整合的十大制约因素,其中前三大制约因素是,语义、语义和语义。语法问题涉及的是格式和结构,而语义关系到信息的理解和信息的一致性的问题。Gartner Group 认为:在信息集成中只有 5%的接口是选择中间件的功能,其余的 95%则是选择应用语义的功能。EAI 提供了将组织现有各类系统集成起来的方法,从根本上来说它提供一个数据格式到另一个数据的翻译过程,基本还是以基于中间件的思想实现异构系统间紧密连接。Web Service 技术是设计和构建机制的突破,在 XML、UDDI、SOAP 等协议的支持下,实现了松散式的服务发现、捆绑、传输消息(服务)到合作系统的机制。但无论是 EAI 还是 Web Service 都没有解决给定交换环境中的不同术语和词汇问题。所以互操作研究的前沿是基于语义的信息互操作 (Semantics-based Information Interoperability),关注解决不相联的词汇、数据定义、术语和全局的视图。基于语义的互操作连接多个系统共享信息、维护共享信息系统资源的自治性、产生一个弹性的信息架构。它的三个利器^[13]是建立在本体上 (Ontology) 之上的语义调节器 (Semantic Mediation)、语义映射 (Semantic Mapping)、上下文关联 (Context Sensitivity)。本体论的研究已经从最初的人工智能研究领域扩展到智能信息集成、协作信息系统、信息检索、电子商务、知识管理等领域^[14],正因如此,本体论成为当前信息领域研究热点之一,也是实现语义互操作的重要工具。

4 元数据技术与公共信息管理整合

当前网络信息的海量性、形式多样性、快速增长性、组织的松散性、动态性、无序性,使得公共信息的管理、检索等工作变得异常复杂。Internet 的发展要求对网络知识空间的信息实施有效的组织、控制,元数据是计算机可识读的关于数字化信息资源的信息,它的管理有利于形成自动化、高质量的网络搜索引擎,有利于网络信息存取和检索。

元数据是关于数据的数据,被用来描述对象特征和属性的数据,是一种编码体系、有效组织网络信息管理工具。元数据在网络信息资源组织管理方面的作用:1)信息描述;2)信息定位;3)信息搜寻;4)信息评估;5)信息选择;6)信息管理。

经过几十年的发展,元数据超越了最初对信息对象的描述和抽象,能够支持对数字信息对象的长期保存,借助于元数据技术已经能够实现对信息资源的评鉴、支持数字权益管理、支持信息使用的政策控制和隐私保护,支持不同系统之间的语义转换进而实现不同系统之间的互操作等。政府领域的元数据是 GILS^[15] (Government Information Locator Service 或 Global Information Locator Service),是一个分布式信息资源利用体系,为公众检索、定位、获取公共联邦信息资源提供方便服务。各政府机构利用 GILS 规定的标准描述信息资源,建立相应的资源目录和检索系统。公众可以通过互联网直接检索这些目录数据,并通过链接直接获得有关数字化资源。GILS 体系的基本构建要素是对具体资源进行描述的元数据,即 GILS 记录 (GILS Locator Record),它用来描述信息资源的内容、位置、服务方式、存取方法等,描述对象主要是来自政府的公共信息资源。在 GILS 体系中,借助于 Z39.50 支持,用户可以一次向 GILS 注册的所有服务器发出检索查询,能够高质量地获取政府公开的信息。

美国从 1995 年起将 GILS 作为政府信息基础设施 (Government Information Infrastructure) 的核心组成部分进行建设,其他如英国、加拿大、澳大利亚等一些信息化水平较高的国家也相继利用 GILS 来建立和推动政府信息广泛共享。英国政府意识到政府仅仅将信息置于 Web 上是不够的,可以通过添加元数据有助于增加信息搜索者定位准确性和利用政府信息的信心,并制定了政府元数据标准 (e-Government Metadata Standard, e-GMS)^[16],用于集成政府部门之间的应用服务

公共领域元数据研究比较活跃的三个方向是:

(1)标准的制定。由于 DC (Dublin Core) 已经成为 Web 资源描述普遍接受的标准,目前各类元数据标准包括各国政府元数据标准的制定大多以 DC 为基础,结合各国或领域的实际状况对现有的 DC 主元素和次级元素进行修改、删除、添加,并制定自己的通过 Lang、Scheme、Sub-Element 规范和约束元数据术语、格式的使用并赋予元素部分含义。

(2)元数据的应用。元数据是下一代互联网 (语义 Web, Semantic Web) 重要基础之一,语义 Web 的两个重要的技术工具是 XML (可扩展标记语言 eXtensible Markup Language) 和 RDF (Resource Description Framework)。RDF 一个目标是在 XML 基础上以标准化互操作的方式规范数据语义,从技术上来讲 RDF 不是一种语言,而是元数据实例的数据模型,或者说是元数据互操作模型。

(3)元数据之间的互操作。包含交换格式、标记格式、元素内容、元素语义、编码规则、数据内容的互操作。目前的机制有直接的映射、通过基于 XML 的 XSLT (扩展格式转换语言) 的模式转化以及 OAI (Open Archive Initiative) 机制。

元数据技术已经成为公共信息整合的重要策略机制,在异质数据库和信息系统集成或各种数据源的互操作中需要各数据源的结构和含义,元数据提供的语义和属性被用来建立系统之间映射的翻译规则^[3],除此之外,元数据在整合过程中被用作内容管理工具。

5 公共信息分类整合

分类是人类认识客观世界的重要方法,它依据一定的标准将具有相同(或相似)特征类型的事物划分成若干个类型或组别,从而实现从复杂向简单的过程转换。政府掌握着种类繁多的公共信息,公共信息分类整合在政府信息化建设中显得非常重要,英美两国的电子政府建设中分别采用了政府分类列表 GCL (Government Category List)^[17] 和共享主题树 (Sharing Subject/Topic Tree)^[18] 的分类方法,对信息进行分类。

英国政府 GCL 是与 E-GMS (e-government metadata standard) 主题元素一道使用的标题分类列表。为适应变革的信息环境,GCL 每年更新 3 次,向 WEBSITE 和门户推荐,建立主题索引、标记电子资源和目录结构。政府公共部门使用 GCL 进行元数据标记使得公民非常容易地发现政府信息。GCL (Version 1.4, April 2003),其顶层有 12 个主题,通过这 12 个顶层主题的导引,可以直达政府部门所提供的服务信息。共享主题树目前已经推出了第五个版本,它包含 24 个顶层政府术语,669 个优选项和 692 个次优选项。

无论是英国的 GCL 还是美国 Sharing Subject/Topic Tree 其根本思想是一致的,即采用一种划分标准,对政府部门的公共信息种类进行划分,给用户在使用公共信息时有一个清晰的信息辨识视角,易于使用,同时也有利于政府公共职能部门信息管理。

另有一种对政府公共信息的管理和发布是按照公民 Life-events 和组织的 Business Situation 来进行的^[19]。这种信息服务的分类的划分体现了政府以人为本的服务理念,使得公民/企业组织根据这种信息分类的视图定义,能快速定位所需服务信息。

6 信息集成(整合)的分类及发展趋势

Mercator 公司研究当前集成整合发展趋势时,将应用集成分为以下三类^[20]:1)面向信息的集成;2)面向过程的集成;3)面向服务的集成。面向信息的集成关注的是应用接口层的转换,应用与系统之间的数据流的流动,为整合的组织提供了低风险、低成本的应用集成的起点,在某些情况下无需改变应用,包括:1)数据复制;2)数据联邦;3)面向接口的集成。面向过程的集成是通过元应用和业务流程为用户提供一种捆绑应用的能力,这一过程是面向信息继承模型顶层的抽象,为我们提供了应用集成问题领域的业务视图和明确的控制流模型。面向服务的集成是动态的应用集成,允许共同业务逻辑大规模的共享。三种应用集成代表着不同方向的不同视图,互补而非排斥。

Andread Koller 将信息集成分为物理集成和逻辑集成^[21],而逻辑集成又分为模式集成、数据集成,模式集成需要比较概念的一致性,概念冲突将被发现并加以解决,结果模式被附加上或被合并,必要时进行重构以改善其完整性、可理解性,消灭冗余。模式集成中的许多工作需要人工辅助,当模式集成完成后,数据集成才开始。这一划分方法仍然是建立在对传统的结构化数据表示及相互之间交流不畅等问题的解决基础之上,尚未考虑到信息技术新的发展环境的变革。

Anup Marwadi 认为集成异质数据源的方法可以分为两类^[3]:1)多数据库系统(或称联邦数据库系统);2)Mediator/Wrapper 系统。多数据库系统为业已存在的数据库提供全局的接口,它解决的是数据库的集成,面向对象的数据模型被广泛地用作全局数据模式,使得本地/全局的数据能够转换、模式翻译和集成,这样的系统易于设计、架构简单,但为了有效地实施查询,用户需要了解虚拟数据库的模式。Mediator 提供中介服务,用于连接数据源和应用程序,实际上是在源数据之上的一层抽象,它的功能是提供集成的信息,无需集成基础数据资源(Base Data Source),它能够访问和检索来自异质资源的相关信息、将检索到的数据抽象和转换成公共的数据表示和语义。这种方法的优势在于它可以解决任意类型的数据、提供共同的接口、将结果转化成抽象、易于理解的中间件格式。Wrapper 接受用户定义的一系列转换规则,将不同的数据格式翻译成特定的模型。

IBM 公司提出了政府信息价值链的思想(Information Value Chain)^[22],用于指导政府决策所需信息在政府部门之间的无缝流动,保证决策时能在正确的时间获得正确的信息。政府公共信息价值链思想的提出,给出了一个全新的认识信息价值的视角,使得我们在进行公共信息整合过程中能从全局的角度或者说能让我们能借助于公共信息生命周期来完整地认识政府公共信息整合流动。长期以来,政府机构评价 IT 系统时,是根据它们如何服务机构的流程需要,而不是回应公众的需要^[23],这就需要政府的管理者跳出现有的思维和工作方式来看待 IT 项目建设,信息价值链为信息整合有效性评价提供了方法论的指导。

综观信息整合技术发展和应用实践,信息集成整合呈现以下的发展趋势。

(1)集成整合从结构化的数据整合向半结构化、无结构化的数据以及三种类型数据混合整合方向发展。

互联网已经成为最大的信息数据库,特别随着基于互联网应用复杂性的不断增加,信息整合集成更显迫切,整合范围已从组织内部发展到组织之间。数据集成意指数据的共享、强化、合并以及不同数据源和应用之间的互换数据,它是 A2A 和互操作的基础^[24],由于互联网中的大量信息表现出半结构化和无结构化的特征,在传统数据库管理系统基础之上建立起来的集成整合方法表示数据语义时使用静态的、严格的结构化数据模式已经不能满足需要。Web 上资源的不断递增,被发布在 Web 上的信息资源不知道最终要进行集成以及不知道要同什么样的信息集成整合,这样坚持语义和模式相关的标准是不可能的。现实世界中,对整合的 Web 资源像传统的信息资源集成那样实施控制更是不现实的,因为分布于 Web 上的资源隶属于不同的组织,这些组织处于一个动态的环境中,信息拥有者从自身的利益出发形成了信息分析、利用检索、定位体系,有着价值最大化的倾向,组织的一个最大偏好是保持自己的信息系统的自治,正由于没有能力改变 Web 资源和其他组织信息资源的构建方式,于是就有了 WRAPPER、中介器、翻译器、映射功能的知识库被用来提供基于 Web 信息的访问、协调视图。

Web 资源的异构性是其整合的最大障碍,资源的异构包括模式异构、数据异构、语义异构。传统的数据库都有自己的数据模式,根据这一模式描述特定的数据,可以很好地定义查询语言,从而很好地实施对数据的操作。半结构化、无结构化的数据通常具有自描述性和动态可变性,它与传统的结构化数据不同在于半结构化、无结构化的数据是先有数据后有模式,这种模式描述是非强制性的,而且处于经常的变化之中,

描述半结构化、无结构化的数据模式的信息就是元数据。

对于半结构化、无结构化的数据的集成整合通过:1)对整合的资源添加结构;2)添加语义,而添加语义方法又可以分为主题映射(Topic Maps)和结构化映射(Structured Maps)^[13]。主题映射是根据正式声明的主题描述信息集合的文档。这些主题与主题关系类型在主体映射文档类型定义(DTD)中声明。主题映射附加的信息通过超链接将主题和资源实例链接起来,结构化映射和主题映射的区别在于其实施和查询能力,它使用关系数据库技术,而被组织的资源仍然存在于 Web 中。在上述基础上可以建立集成整合的一致视图,这里可以使用多种技术(元数据技术、XML RDF、WRAPPER、MEDIATOR 等)。

(2)信息整合正从数据结构或语法集成整合向信息语义、知识整合方向发展。

我们知道在 Web 资源的组织管理中使用的常见的两种技术是搜索引擎(SE)以及信息中心索引技术,搜索引擎会带来大量的毫不相关的信息,其原因在于它的关键字检索,这种机制将输入的关键字仅仅视作毫无含义的一般字符;而中心索引方法也随着资源的不断丰富以及不断的变化而变得越来越臃肿^[24]。

通过 RDF、XML 和其他的方案,Web 能够在相关联的信息中嵌入语义信息,或者通过信息抽取技术、本体技术或其他技术在被集成的资源之上建立一层抽象的语义层,有了抽象的语义层,可以开发查全率和查准率更好的 SE,可以在不同的组织、系统之间实现知识的共享重用。

(3)信息集成整合系统之间由紧耦合向松散耦合方向发展。

W3C 规范应用集成框架目标是^[3]:1)集成架构应能使服务的开发沿着更广阔的范围进行;2)框架必须可靠和稳定。客观环境的变化,使得一个组织的用户、伙伴处于不断动态变化之中,在变化的环境中,与本组织进行信息集成整合的系统也在发生变化;传统的建立在紧耦合基础之上的信息集成整合方法越来越不能满足要求,如 CORBA 以及 COM 或 COM+ 技术等由于自身的特征使得这些技术仅适合组织的内部使用。

Web Service 是一种分布式的计算技术,通过 URI 识别的软件系统,其公共接口和绑定是利用 XML 来定义和描述的,能够被其他软件系统发现,同时这些系统可以在 Internet 协议上使用基于 XML 消息与 Web Service 按照其描述的方式进行交互。

它是位于 Web 上的松散耦合、动态定位的软件组件,这些组件将大的软件系统剖分成更小的逻辑模块或共享服务,这些服务可以位于不同的系统,可以用各种技术来实现,Web Service 已经成为互联网世界中面向服务的关键,因为它能与未知的、不相关的应用以简单的方式进行相互之间的交互。Web Service 的关键技术有 SOAP、UDDI、WDSL、XML,不像传统的端到端以及硬编码的开发接口方法,它使松散耦合具体化,应用服务注册、服务和集成动态化,与此同时 Web Service 的另一个优势是,它可以在传统的 EAI 解决方案之上进行实施,通过 XML 等技术将原有的服务进行封装成 Web 服务组件,这样组织没有必要替换现有的系统,促进了系统的升级、保护了原有的投资,降低了成本和风险,使得整合可以按照渐进的方式来进行。这一点对公共信息整合非常有启发,由于公共信息所服务的对象的复杂性、责任的重大性以及运行的连续性,决定着公共信息整合绝对不能是跳跃性的,因而 Web Service 技术体系应用也就成了公共信息整合的重要策

略。

(4)信息整合中的数据正从分散走向集中。

一种是数据的物理集中,一种是数据的逻辑集中。数据的逻辑集中表现为:(1)为保持资源的自治性,对集成整合的各种非结构化资源通过诸如:E-R 模型等进行添加信息语义或结构,再利用一种机制对包括信息资源之间的关系实现集中管理,资源之间仍然通过超链接进行访问。(2)通过 Mediator/Agent/Adapter 结构,将整合的数据源模式(Schema)注册于 Mediator 中,在 Mediator 中建立通过模式映射建立共同的数据视图——视图集中,在 Agent 和 Adapter 的帮助下,实现于与各数据源的连接访问。

结束语 面对不断更新的信息技术、面对公众不断提高的新的信息需求,信息系统运行环境已经发生了深刻的变化;公共信息、公共信息系统唯有进行整合才能为公众提供高质量的信息服务。诚然公共领域信息服务所面临的环境是复杂的,要解决的问题很多,包括组织、制度、标准和法律等,加强公共信息整合机制策略研究,可以为我国电子政府的发展建设提供有益的参考,为电子政府的具体重大信息化项目规划、建设提供科学建议,解决现行我国 e-GOV 建设过程中偏离发展主轴的部分问题。由于我国各地区、各政府部门信息化建设的均衡性,公共信息整合问题的研究可以为具有后发优势的政府部门公共信息领域信息化建设提供跨越式发展的理论指导,促进我国电子政务建设的均衡发展,对逐步构建一个服务型、责任型、法制型开放型政府,为公众提供一体化、集成化、透明化全方位的快速公共信息服务,建立一种新型的政府与公民、政府与企业、政府与政府之间良好社会互动关系具有重要的理论和实践意义。

参考文献

- 1 Wimmer M A. A European perspective towards online one-stop Government : the e-GOV project. *Electronic commerce Research and Application*, 2002(1): 92~103
- 2 A comprehensive assessment of public Information Dissemination. http://www.nclis.gov/govt/assess/assess_voll.pdf
- 3 Marwadi A. *Ontological Semantic Integration Model*: [Master thesis]. University of Missouri-Kansas city, 2002
- 4 Industry advisory council. *Interoperability strategy: concept, challenges, and recommendations*. <http://www.iaconline.org/documents-presentations/2003/030403-interoperability-strategy-concepts-challenge-and-recommendations.pdf>
- 5 e-Government Interoperability Framework <http://www.govtalk.gov.uk/documents/vol2-20030106.pdf>
- 6 A New Zealand E-Government Interoperability Framework. www.e-government.govt.nz/docs/e-gif-v-0-9/e-gif-v-0-9.pdf
- 7 Moen W E. Introduction to GILS. In: *The third annual GILS conf.* March 2001
- 8 Layne K, Lee J. Developing fully functional E-government : A four stage model. *Government information quarterly*, 2001, 18: 122~136
- 9 Moen W E. Mapping the Interoperability Landscape for networked information Retrieval. <http://www.unt.edu/wmoeu/publications/mapinteropjcdlfinal.pdf>
- 10 Miller P. Interoperability what is it and why should I want it? <http://www.ariadne.ac.uk/issue24/-Interoperability/intro.html>
- 11 Friesen N. Semantic Interoperability and communities of practice. www.educationau.edu.au/globalsummit/papers/nfriesen.htm
- 12 Z39. 50. <http://www.oclc.org/content/45/pdf/z3950-handbook-paper.pdf>
- 13 Integration's dirty little secret: It's a matter of semantics. www.modulant.com/pdfs/dirty-little-secret.pdf
- 14 Fensel D, van Harmelen F. OIL: An Ontology Infrastructure for the Semantic Web. *IEEE Intelligent Systems*, 2001, 16(2)
- 15 Application profile for the government information locator service (gils). www.itl.nist.gov/fipspubs/192-a.pdf Information Technology Laboratory AUGUST 1997

(下转第 15 页)

杂交,产同样数目的孩子粒子,并用孩子粒子代替父母粒子,以保持群的粒子数日不变。孩子粒子的位置由父母粒子的位置的加权和计算,即:

$$child_1(x) = p \times parent_1(x) + (1-p) \times parent_2(x) \quad (5)$$

$$child_2(x) = p \times parent_2(x) + (1-p) \times parent_1(x) \quad (6)$$

其中 x 是 D 维的位置向量,而 $child_k(x)$ 和 $parent_k(x)$, $k=1,2$ 分别指明是孩子粒子还是父母粒子的位置; p 是 D 维均匀分布的随机数向量, p 的每个分量都在 $[0,1]$ 取值,' \times ' 表示向量分量对应相乘。孩子粒子的速度分别由下面的公式得到:

$$child_1(v) = \frac{parent_1(v) + parent_2(v)}{|parent_1(v) + parent_2(v)|} |parent_1(v)| \quad (7)$$

$$child_2(v) = \frac{parent_1(v) + parent_2(v)}{|parent_1(v) + parent_2(v)|} |parent_2(v)| \quad (8)$$

其中 v 是 D 维的速度向量,而 $child_k(v)$ 和 $parent_k(v)$, $k=1,2$ 分别指明是孩子粒子还是父母粒子的速度。杂交粒子群优化算法引入了较多的待调整参数,对使用者的经验有一定要求。

表 2 CPSO、IWPSO、CRPSO 和 MPSO 算法数值仿真结果

	全局最小值点平均值	全局最小值平均值
CPSO	(0.723e-006, 0.418e-006)	5.018e-010
IWPSO	(0.239e-005, 0.393e-006)	8.175e-009
CRPSO	(0.00130518, 0.00086676)	0.00346642
MPSO	(0.00022500, 0.00036320)	0.00622002

文[8]进一步把变异运算引入粒子群优化算法中,提出了带高斯变异的粒子群优化算法,变异运算依据变异概率选取指定数量的粒子按高斯变异算子进行变异,用变异后的粒子代替原粒子,即:

$$mutation(x) = x * (1 + Gaussian(\sigma)) \quad (9)$$

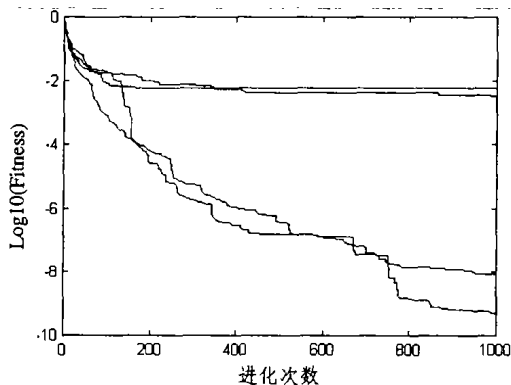


图 2 CPSO、IWPSO、CRPSO 和 MPSO 算法进化曲线比较

图 2 是 CPSO、IWPSO、CRPSO 和 MPSO 算法收敛曲线比较,从上往下的四条曲线依次对应的是 MPSO 算法、CRPSO 算法、IWPSO 算法和 CPSO 算法。表 2 是这四种算法数值仿真结果。从图 2 和表 2 可知混沌粒子群优化算法优于杂交粒子群优化算法和带高斯变异的粒子群优化算法,而略好于

带惯性因子的粒子群优化算法。

结论 本文把混沌优化思想引入到粒子群优化算法中,给出了混沌粒子群优化算法。其基本思想是在粒子群的每次进化过程中,首先对粒子群体中的最优粒子进行混沌寻优,然后随机地从粒子群体中选择一个粒子用混沌寻优的结果来替换。通过这种处理使得粒子群优化算法摆脱局部极值点的能力得到改善,提高了算法的收敛速度和精度。计算机仿真结果表明混沌粒子群优化算法的收敛性能明显优于粒子群优化算法,亦优于杂交粒子群优化算法和带高斯变异的粒子群优化算法,而略好于带惯性因子的粒子群优化算法。

参考文献

- Kennedy J, et al. Particle swarm optimization. In: IEEE Int'l Conf. on Neural Networks. Perth, Australia, 1995. 1942~1948
- Eberhart R, Kennedy J. A new optimizer using particle swarm theory. In: Proc. of the sixth intl. symposium on Micro Machine and Human Science, Nagoya, Japan, 1995. 39~43
- Shi Y, et al. A modified particle swarm optimizer [C]. In: IEEE World Congress on Computational Intelligence, 1998. 69~73
- Shi Y, Eberhart R C. Fuzzy Adaptive particle swarm optimization [C]. In: Proc. of the Congress on Evolutionary Computation, Seoul Korea, 2001
- Clerc M. The swarm and the Queen: Towards a deterministic and adaptive particle swarm optimization [C]. In: Proc. of the Congress of Evolutionary Computation, 1999. 1951~1957
- Angeline P J. Evolutionary optimization versus particle swarm optimization: Philosophy and performance differences [C]. In: Evolutionary programming VII, 1998. 601~610
- Lovbjerg M, Rasmussen T K, Krink T. Hybrid particle swarm optimization with breeding and subpopulations [C]. In: Proc. of the third Genetic and Evolutionary computation conf. San Francisco, USA, 2001
- Higashi N, Iba H. Particle swarm optimization with Gaussian mutation [C]. In: Proc. of the Congress on Evolutionary Computation, 2003. 72~79
- Van den Bergh F, Engelbrecht A P. Training product unit networks using cooperative particle swarm optimizers [C]. In: Proc. of the third Genetic and Evolutionary computation conf. San Francisco, USA, 2001
- Van den Bergh F, Engelbrecht A P. Effects of swarm size cooperative particle swarm optimizers [C]. In: Proc. of the third Genetic and Evolutionary computation conf. San Francisco, USA, 2001
- Kennedy J, Eberhart R. Discrete binary version of the particle swarm algorithm [C]. IEEE Int'l Conf. on computational Cybernetics and Simulation, 1997. 4104~4108
- Clerc M. Discrete particle swarm optimization illustrated by the traveling salesman problem. <http://www.mauriceclerc.net>, 2000
- Ciuprina G, Ioan D, Munteanu I. Use of intelligent-particle swarm optimization in electromagnetics. IEEE Trans. on Magnetics, 2002, 38(2): 1037~1040
- Clerc M, Kennedy J. The particle swarm-explosion, stability, and convergence in a multidimensional complex space. IEEE Trans. on Evolutionary Computation, 2002, 6(1): 58~73
- 王东升, 曹磊. 混沌、分形及其应用 [M]. 合肥: 中国科学技术大学出版社, 1995
- 李兵, 蒋慰孙. 混沌优化方法及其应用 [J]. 控制理论与应用, 1997(4): 613~615

(上接第 12 页)

- e-government metadata Framework. [http://www.govtalk.gov.uk/documents/uk metadata framework vol202001-05.pdf](http://www.govtalk.gov.uk/documents/uk%20metadata%20framework%20vol202001-05.pdf)
- GCL (Government Category List) <http://www.govtalk.gov.uk/schemasstandards/gcl.asp>
- Mullen A. GILS metadata initiatives at the state lev-el. Government Information Quarterly, 2001, 18: 167~180
- Jordan, readiness for e-government. <http://www.surf-as.org/papers/e-gov-english.pdf>
- Linthcum D S. Mercator: Next generation Application integration From Infomation, or Process, to Service Mercator software, Inc,

2002

- Koeller A. Integration of Heterogeneous Database Discovery of Meta-Information and Maintenance of Schema-restructuring views: [PHD thesis]. Worcester polytechnic Institute, 2001
- IBM Corporation. Optimization of information to improve decision making in government:- the information value chain way. <http://www.ibm.com/services/files/ibv-infovaluech-ain.pdf>
- E-Government Strategy. <http://www.cio.gov/documents/egovreport.pdf>
- Reese W D. An Investigation of Techniques for Integrating Web-accessible Data: [PHD thesis]. University of Colorado, 2001