

信息资源整合技术初探

彭洁 赵辉 齐娜

(中国科学技术信息研究所,北京 100038)

摘要:我国信息化发展取得了积极进展,信息技术应用与国民经济和社会发展的融合日益增强。随着各种信息系统建设的不断完善,信息化建设进入了信息资源整合阶段。本文在深入研究的基础上,对信息资源整合技术的概念、发展历程、发展环境和技术框架体系、发展趋势等进行阐述。

关键词:信息资源整合;信息资源;整合技术;信息技术

中图分类号:G250.73 **文献标识码:**A **DOI:** 10.3772/j.issn.1674-1544.2008.01.008

Research on Information Resource Integration Technology

Peng Jie, Zhao Hui, Qi Na

(Institute of Scientific & Technical Information of China, Beijing 100038)

Abstract: Information integration is an important component in today's information industry. The paper defines the information integration, introduces the history, the technology bases, the framework and the trend of integration technology development.

Keywords: information resource integration, information resource, integration technology, information technology

我国信息化从上世纪90年代早期开始起步,发展到今天,面临着信息系统缺乏统一规划和信息资源标准不一致等问题。一个集团性企业总部有时甚至会用到多达100个到200个信息系统,既有单机版的,也有网络版的。不同的应用系统建立在不同的硬件和操作平台上,有Unix系统的,也有Windows系统的;不同的应用系统是由不同的系统集成商使用不同的语言和开发工具开发出来的,有用ORACLE的,也有用SQL的;数据结构也不一致,例如同样是描述姓名属性的字段,在数据库中可能会有不同的定义方式。在实际使用时,这使得同一业务系统的不同模块之间以及不同业务系统之间的信息共享十分复杂,业务运作不流畅,需要过多的人工介入,

致使不同的应用系统之间形成了彼此隔离的所谓“信息孤岛”(信息资源分散和异构)。彻底消除信息孤岛已是当前信息化建设的瓶颈,也是必须解决的问题。然而,解决这一问题的关键是开发和利用信息资源整合技术,整合现有的信息资源。

1 信息资源整合技术的概念

美国信息资源管理委员会资深顾问F.W. Horton博士与D.A. Marchand合著的《信息趋势:从你的信息资源中获利》(Infotrends: Profiting from Your Information Resources, 1986)一书中认为,对于整个社会和国家来说,信息资源包括如

下4方面内容:具有与信息相关技能的人才;信息技术中的硬件和软件;信息机构,如图书馆、计算中心、通信中心和信息中心等;信息处理服务提供者。对于商业性机构,他们提出了“信息资产”的概念,认为信息资产包括:公司拥有的数据和文献资料;公司拥有的“know-how”,包括知识产权(专利、著作权)和个人(专家)的经验;公司拥有的竞争性信息,即有关其竞争对手和商业环境的信息。

此外,波拉特认为,信息资源包括信息的处理、操作、传递中所用到的一切人、机械、产品、服务^[1]。这些都是信息活动不可缺少的因素。这些资源若不加以控制、利用,那么扩大信息活动是不可能的。德国信息管理专家 K. A. Stroetmann (1992年)认为,信息资源包括3个组成部分:信息内容、信息系统和信息基础结构^[2]。我国学者乌家培认为,对信息资源可以有两种理解:一种是狭义的理解,即指信息内容本身。另一种是广义的理解,指的是除信息内容本身以外,还包括与其紧密相连的信息设备、信息人员、信息系统、信息网络等。狭义的信息资源实际上还包括信息载体,因为信息内容不能离开信息载体而独立存在^[3]。

不难看出,广义的信息资源强调了集成性,是信息活动中涉及各要素的集合,而狭义的信息资源则比较简单,只涉及信息内容本身。综上分析,笔者给出了这样的定义:信息资源是指由人类活动各个领域所产生的、能够在计算机系统存储、检索和管理的各种信息、系统、网络等的集合。与物质资源相比,信息资源具有如下特性:

(1) 共享性与选择性:物质资源和能源资源的利用表现为占有和消耗,而信息资源的利用不存在竞争关系,各利用者可以同等程度地共享某一份信息资源^[4]。共享性是信息资源的一种本质属性。在信息资源的使用上,不同的使用者根据自己的意图把信息资源用于不同方面,从而产生不同的效果。

(2) 稀缺性与时效性:信息资源的稀缺性反映在两个方面,一是在既定的时间和空间里,某一特定的经济活动行为者由于人力、物力、财力等因素的限制,其信息资源的拥有量总是有限的^[4]。二是

信息资源的总效用将随着使用次数的增多而逐渐衰减。信息资源特别是与决策相关的信息都要求具有高度的时间敏感性,即要求及时性和适时性。

(3) 分散性与交叉性:分散性主要是指信息资源主要分布在不同实体中或不同载体上,而信息之间互相关联和交叉,即反映出了交叉性的特点。

(4) 剧增性与多样性:随着信息技术和现代科技的发展,各种信息资源急剧增加。特别是网络信息资源以其剧增性引起人们的重视。从载体形式和内容看,信息资源也呈现出多样性,如数据、文献、多媒体等,而内容更是包括了各个行业,不同学科。

(5) 不稳定性与价值差异性:不稳定性主要反映在互联网信息资源方面,信息的产生和消除具有不确定性。同时,互联网信息由于缺少相应的控制,信息质量和价值层次不齐,反映出价值差异性。

那么,什么是整合信息资源和信息整合技术呢?简单地说,整合信息资源就是用统一接口、统一标准访问在异构网络、异构数据、异构平台下的信息资源并进行深层次开发利用的活动,而实现这一功能的技术则称为信息资源整合技术。实际上,整合信息资源是一项复杂的系统工程,单纯的一项技术是不能解决信息资源分散和异构问题的,而是与整合信息资源相关的一系列技术一起解决信息资源的分散和异构问题。因此,信息整合技术是一组技术,是一个技术群。

2 信息资源整合技术的发展历程

就世界的发展而言,信息资源整合技术是伴随着信息技术和信息化应用的发展而逐渐发展起来的。对比信息技术的发展,从上世纪六七十年代的计算机技术大发展时代,到1985年之后的图形化时代,20世纪90年代开始的企业信息化时代,以及20世纪90年代中后期到现在的互联网时代,伴随着信息大爆炸,信息资源大规模增加,但信息的利用程度并没有提高,因此,整合信息资源势必得到日益广泛的应用。我国的情况

是信息化滞后西方国家 5~10 年,因此信息资源整合大体上在 2000 年才启动,到目前进入整合提升阶段^[5]。所以,我国信息资源整合技术也是在信息化建设过程中面临整合问题时提出,并在应用中发展成熟起来的。

信息资源整合技术发展大体上经历了 3 个阶段(图 1)^[6]:

(1) 萌芽阶段(1980-1995): 上世纪 80 年代中后期到 90 年代中期,提出了许多信息资源的整合技术,并开始逐渐产品化,如 Web Services、中间件、搜索引擎技术、博客技术、群决策技术等。这个阶段,在这些技术创生时,人们并没有意识到它将被用于信息资源整合,但随着技术的逐渐成熟和广泛应用,信息资源整合的属性被强化,跨平台问题的提出,使得这些技术发挥了最大价值^[7]。

(2) 大发展阶段(1996-2000): 信息资源整合技术的作用逐渐发挥出来,新的产品不断创新,大的厂商逐渐形成了一个市场,技术在进一

步发展。我国政府将信息资源整合纳入了议程,并在部分地区开始试点。总体来看,这一阶段,信息资源整合技术获得大范围的应用和拓展,并在应用中不断完善和改进^[8]。

(3) 应用深化阶段(2001-): 进入新世纪以来,信息资源整合技术的价值已被用户广泛认知,应用领域和范围越来越广阔,相关技术也越来越成熟。预计在不久的将来,各种实际应用将使各种技术有机地整合起来,实现价值最大化、融合化,在技术上将有所突破。

3 信息资源整合技术的技术基础

信息资源整合技术发展到现在水平,是以信息技术的发展为依托的。具体地讲,这些信息技术发展的特点是^[9-11]:

(1) 计算机用集成电路的集成度迅速从小规模发展到大规模、超大规模的水平,计算和存储能力大幅度提高,为在硬件上集成信息资源提

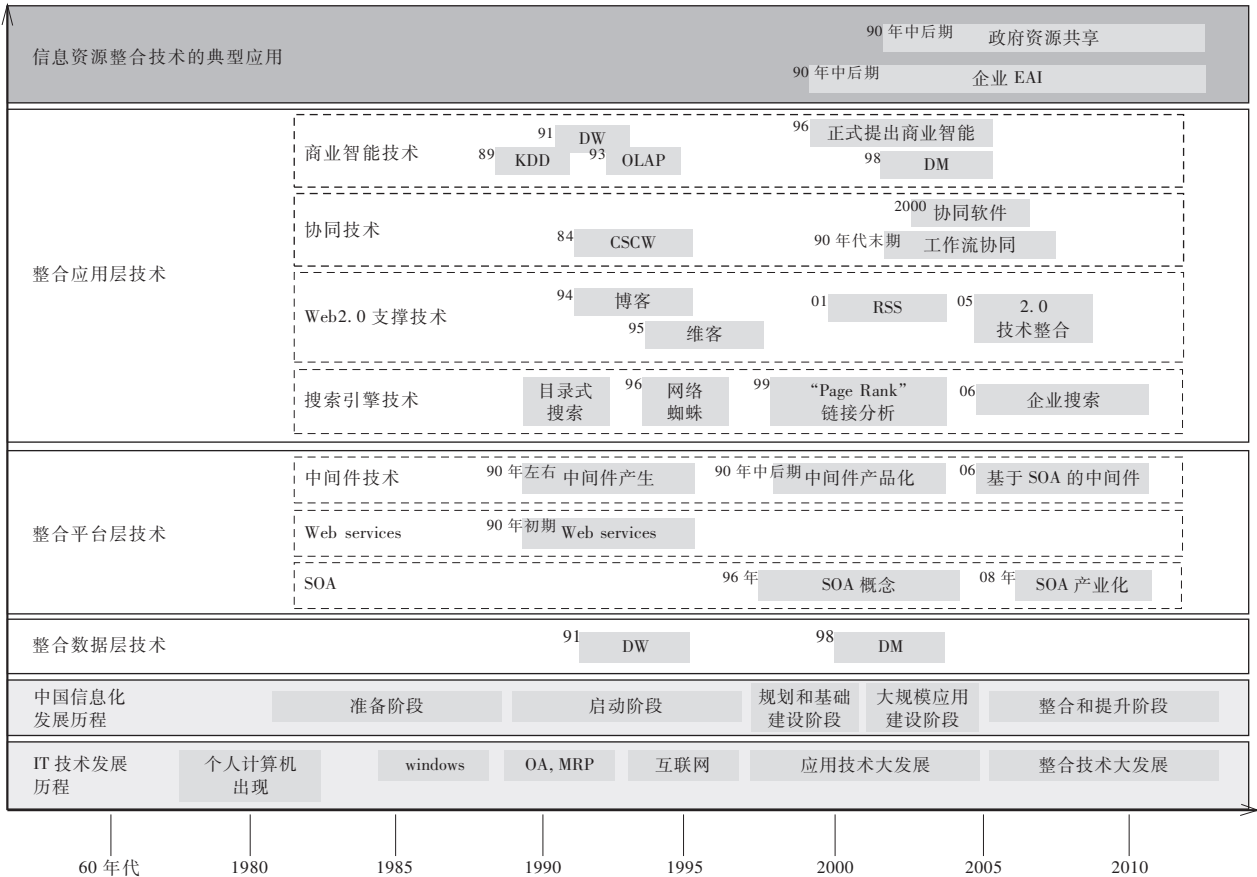


图 1 信息资源整合技术发展历程

供了可能。

(2) 集数据管理、对象管理和知识管理为一体的第三代数据库系统的产生以及数据仓库概念的提出和产品化,为进一步实现信息资源的管理提供了软件支持。

(3) Web 互联网和快速以太网可使网络传输速度达到 100Mbps,为高速传送信息提供了传递通道。

(4) 各种具有不同功能、能够适应不同环境的服务器不断地出现,多种多样的服务器的发展可满足用户多方面利用信息的需求。

(5) NAS 网络附加存储、SAN 存储区域网以及存储网络系统等大规模集群存储和对等存储技术的发展,提供了更大的信息贮备空间。

(6) 移动终端的发展可做到方便快捷地采集和获取信息。

(7) 信息资源交互标准的制定,为信息整合、交流与共享奠定了基础^[12,13]。

4 信息资源整合技术框架

按照信息资源整合技术发挥作用的特点,可将其分为标准和规范层、整合基础网络层、整合数据层、整合平台层和整合应用层,如图 2 所示。

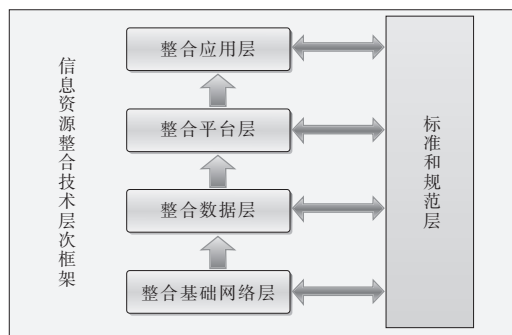


图 2 信息资源整合技术的层次

现根据这些层次,列举出相关的技术,如图 3 所示。

4.1 标准和规范层

标准和规范层是指支撑信息资源整合技术的各种标准和规范,包括信息技术基础标准体

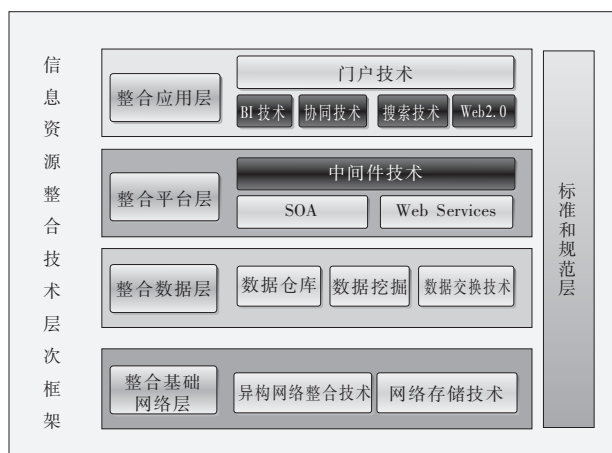


图 3 信息资源整合技术框架图

系、信息资源标准体系、网络基础设施标准体系、信息安全标准体系、应用标准体系及管理标准体系。

4.2 整合基础网络层

整合基础网络层是指在异构网络整合方面涉及的技术以及实现网络环境下信息资源集中存储的技术,主要有:

(1) 异构网络的协议转换技术:一方面是通过软件的方式进行协议转换,即通过建立一种映射机制,将某一协议的收发信息(事件)序列映射成另一协议的收发信息(事件)序列;另一方面是通过硬件的专门协议转换网关设备(也有产品称为网络控制器)进行实时的异构网络协议转化,实现对客户完全透明的信息资源交换和整合过程。

(2) 网络存储管理技术:比较常见的技术主要有 NAS(网络附加存储)和 SAN(存储区域网络)两种。此外,近年还开发了一种虚拟存储技术,可以简化存储管理的复杂性,降低存储管理和运行成本。

4.3 整合数据层

整合数据层是指数据分布存储的数据仓库构建以及数据跨平台的交换,主要技术包括:

(1) 数据仓库技术:解决数据的逻辑存储,并作为商业决策支持数据的物理实现模型,存储和管理用户战略决策所需的所有信息。

(2) 数据挖掘技术:指按照预设的规则对数

数据库和数据仓库中已有的数据进行信息开采、挖掘和分析,从中发现和抽取非直接的模式或面向需求的知识,利用这些模式和知识为决策者提供决策依据,实现数据从细粒度到粗粒度的知识整合。

(3) 数据交换技术:目前主要有两种实现方式,一是基于数据总线的方式,通过传输总线,连接了分布异构系统中的各个构件,也就是通过总线来调度数据交换进程;二是基于事件的驱动方式,通过事件任务的触发,利用任务的调度技术,实现数据转换和数据交换存储。

4.4 整合平台层

整合平台层分两个层次:一是着重解决跨平台整合的架构技术和理念,以及标准和相关技术,如 SOA、Web Services 等;二是重点关注解决具体的“信息孤岛”问题,如中间件技术。

(1) SOA 技术:是一种软件系统架构技术。它以客户为中心,将企业或政府机构的信息资源转变成一个个有机结合的组件,按照客户的需求整合相关的组件资源,从而让企业或政府机构真正实现灵活整合,保证了应用系统的灵活性^[14-17]。

(2) Web Services 技术:提供了一种分布的、与平台无关的应用程序机制,能够在微软、Sun、IBM 等公司所开发的不同平台之上的应用程序间协同工作,实现异构信息的共享、交互与整合利用。Web Services 实现应用程序在跨平台和跨网络时良好通信,例如,应用程序集成、B to B 集成、代码和数据重用以及通过 Web 进行客户端和服务器的通信等。

(3) 中间件技术:指处于应用软件和系统软件之间的一类软件,是独立于硬件或数据库厂商(处于其产品的中间,实现其互连)的一类软件,是客户方与服务方之间的连接件,解决信息的分布异构问题^[18]。

4.5 整合应用层

整合应用层直接面对终端用户,在用户的工作流程、协同、交互过程中完成信息资源的集成与整合以及深度挖掘和利用。主要的技术包括:

(1) 商业智能技术:赋予机构信息整合与分

析的优异能力,并迅速汇总机构内外部的数据,将其转化为有效的决策信息。它在商业企业已经得到较大范围的应用^[19-23]。

(2) 协同技术:所谓协同,就是指协调两个或者两个以上的不同资源或者个体,使其协同一致地完成某一目标的能力。协同技术就是实现这种能力的一系列技术,最典型的产品形态并为用户普遍接受的就是协同软件^[24]。

(3) 搜索技术:实现对信息资源的有效获取。狭义上讲,搜索技术一般指以 Yahoo、Google、百度为代表的网页搜索引擎技术,即通过网络蜘蛛技术抓取互联网上的海量网页,通过链接分析、点击分析等技术对网页进行多种索引排序的技术。广义上讲,搜索技术应该是一套强大的信息处理工具包,能帮助人们快速准确地获取信息,处理对象包括传统意义上的数据库,也包括网页、文本文档、邮件、图片等多格式的非结构化信息,而这些信息的来源有互联网、数据库、邮件系统、文件系统,甚至 OA、ERP 等多种业务系统^[25]。

(4) Web2.0 支撑技术:是一套采用开放式开发架构,可极大地丰富互联网应用及信息共享程度的技术体系,是指支撑 Web2.0 模式的一类技术的集合,如 Blog、RSS、SNS、Wiki、Tags 等,最终将允许用户进行端到端的信息创生、发布与共享^[26]。

(5) 门户技术:是帮助用户将机构内所有的系统资源整合成一个入口,用户可以访问所有的信息、文档、流程和业务应用,并将所有表现内容门户化的技术,通常会与搜索技术、中间件技术、协同技术、商业智能技术、Web2.0 等多种技术综合使用,对信息资源进行从内容到门户的全面整合,形成信息资源整合的最终界面^[27]。

5 信息资源整合技术发展趋势

信息资源技术的发展,一直存在着两条主线。一条是围绕互联网信息资源的共享和利用而发展起来的,如搜索技术、RSS 技术、博客技术、维客技术、SNS 技术等。但目前支持互联网的信息资源由于具有公众性、开放性的特点,涉及到企业内部或机构内部的信息是有限的。另一条是围绕解决企业或机构内部信息资源整合而发展

起来的。如中间件技术、SOA、数据交换技术、数据仓库、数据挖掘、CT(协同技术)以及 Web Services 技术等。近年来,基于互联网的技术逐渐在以 Web2.0 为代表的新一代互联网技术中被整合利用,并以 Web2.0 网站的形式为大众提供服务。而企业内部的信息资源整合技术通过 EAI、资源共享平台被整合利用,中间技术融合了 SOA、Web Services 技术,目前最新的中间件产品都是基于 SOA 技术构建的。

就我国当前情况来看,随着信息技术的发展以及企业应用层次的提高,企业竞争形式发生了改变,由企业之间的竞争逐渐演变为基于供应链的全球化竞争。政府也愈来愈开放,政府的服务功能在强化,政府与企业、政府与社会的信息交互需求愈加强烈,信息整合已经不仅仅局限于企业或机构内部,或者有限的企业、机构或个人,必然会涉及到企业内部信息资源、互联网信息资源、政府信息资源、个人信息资源的全面整合。这就要求各种信息资源整合技术的综合应用。一方面通过技术间的组合实现,另一方面作为技术提供方的厂商,为获取更多的客户,提供一站式的服务,技术链条呈现纵向整合趋势。在政府信息共享、企业应用集成等领域,对信息资源整合技术的整合应用已经成为现实。对于政府信息共享来说,不仅需要整合政府部门内部的信息资源,还需要整合互联网的信息资源,这样才能发挥其他机构或企业信息资源的作用,同时也需要多样化的互联网技术为用户提供便利的服务。

目前,信息资源整合技术的发展呈现出四大趋势:

(1) 技术融合化。信息资源整合技术融合化的趋势主要表现在以下几个方面:一是企业级内部信息资源整合技术与互联网信息资源整合技术的融合,如搜索技术在行业企业的应用;协同技术与博客技术、即时通信技术的融合;政府信息共享需要对 Web2.0 技术和 EAI 技术的融合。二是互联网信息资源整合技术之间的融合,如 Web2.0 平台技术对 RSS、搜索技术、博客技术、SNS、维客技术的融合,博客技术与搜索技术的融合,如博客搜索等。三是行业应用信息资源

整合技术的融合。如中间件技术与 SOA 的融合。整体来看,技术属性的互相渗透和关联已经成为一种趋势。

(2) 应用协同化。从信息化最新需求的角度看,原来相对固化的管理需求正向组织之间动态的协同管理需求转化,固定的流程管理也在向动态的流程协同转化,动态要求系统充分“柔性”,能有效支持协同工作的需求。同时,信息化由原来垂直的 IT 系统建设到扁平 IT 系统建设,系统应用之间协同的需求日益迫切。因此,协同将成为信息资源整合的基本需求之一,信息资源整合技术也将向满足协同应用需求的方向发展。

(3) 服务智能化。以商业智能为代表的服务智能化趋势日益凸显。一方面商业智能的发展逐步与协同技术、中间件技术、搜索技术等相融合。另一方面其他信息资源整合技术的传统应用领域,在整合信息资源的基础上进行数据挖掘、分析和展现,已经成为信息资源整合的基本服务。比较典型的是,在搜索技术最新发展中,已经吸收和引入了商业智能的数据分析挖掘的元素。在最新的企业级搜索技术和产品中,结合种类齐全的抓取器、内容和数据分析挖掘工具、分类和聚类方法、可视化显示技术、智能化和个性化的服务,对搜索的结果进行信息分析并支持业务系统的决策分析,呈现出服务智能化的特征。

(4) 系统平台化。信息资源整合技术的发展就是为了解决异构数据、跨平台、跨网络的信息整合和利用问题。从客户容易管理和维护的角度来看,如果系统比较刚性,那么,用户在使用资源时,其切换成本将很高。因此,在以 SOA 技术为核心发展起来的平台化、组件化理念逐渐被厂商和用户接受。如中间件技术、协同技术、BI 技术都逐渐引入平台化的理念,呈现出平台化的趋势。总体来看,平台化成为整个应用软件的趋势,信息资源整合技术也不例外。用基础公共技术开发出平台应用软件,以组件的形式应用于个性化的系统解决方案中,一方面将会降低开发者的难度,同时也有利于降低用户的管理、维护和更新成本。信息资源整合技术面对的元数据源可能处在不同的环境中,但对应用者来说是透明的。从这个意义上讲,平台化将提供解决这个问题的有效

方法。因此,以平台为核心进行融合已经成为未来发展方向。

参考文献

- [1] 赖茂生. 信息资源管理教程[M]. 北京:清华大学出版社,2006.
- [2] 赖茂生. 信息资源管理概述[EB/OL]. http://irm.im.pku.edu.cn/old_website/download/information_resource_management.pdf.
- [3] 薛红英,张秀平. 我国信息经济学研究综述[J]. 经济论坛,2004(8):153.
- [4] 姜明辉,杨海波,王喆. 技术创新信息资源的特征与功能[J]. 中国科技成果,2007(3).
- [5] 刘振兴. 我国与信息化强国的差距在哪? [N]. 中国计算机报,2007-08-07.
- [6] 彭洁,赵辉,杨青峰. 信息资源整合技术框架构建[J]. 江苏科技大学学报(社会科学版),2007,7(1):46-48.
- [7] 赵辉,杨青峰. 用技术整合信息资源[N]. 计算机世界报,2007-07-16.
- [8] 高复先. 中国信息化趋势报告(三十四)信息资源整合技术报告(1)[J]. 中国信息界,2005(20).
- [9] 周卫娟,宋晓群,吴相效. 江苏省国土资源基础数据整合技术研究与实践[EB/OL]. http://www.lm.cn/bookscollection/magazines/maginfomatization/2007maginfomatization/xinxihua200705/200712/t20071227_183108.htm.
- [10] 赵辉,杨青峰. 用技术整合信息资源破除信息孤岛难题[EB/OL]. <http://www.cioage.com/art/200707/47566.htm>.
- [11] 关于对政府信息资源整合有关概念的理解和认识[EB/OL]. <http://www.hbst.gov.cn/info.jsp?id=1240>.
- [12] 整合信息资源[EB/OL]. http://media.ccidnet.com/art/2615/20060903/889829_1.html.
- [13] 蔡文海. 企业信息资源整合与共享之路[EB/OL].

<http://articles.e-works.net.cn/erp/article9880.htm>.

- [14] Thomas ERL. SOA 概念、技术与设计[M]. 北京:机械工业出版社,2007.
- [15] 什么是 SOA? [EB/OL]. http://tech.ccidnet.com/art/287/20060109/409821_1.html.
- [16] 什么是 SOA? [EB/OL]. <http://tech.it168.com/m/2007-09-12/200709121313829.shtml>.
- [17] 什么是 SOA? [EB/OL]. <http://www.enet.com.cn/article/2007/0404/A20070404523706.shtml>.
- [18] 中间件技术的概念和分类[EB/OL]. <http://forums.cweek.com.cn/archiver/tid-104152.html>.
- [19] 谢榕. 基于数据仓库的决策支持系统框架[J]. 系统工程理论与实践,2000,4:27-30.
- [20] 楼伟进,楼伟忠. 数据仓库与知识发现[J]. 计算机工程与应用,2000,36(10):111-113.
- [21] 王珊等. 数据仓库技术与联机分析处理[M]. 北京:科学出版社,1998.
- [22] 黎锁平,李娟. 基于数据仓库的 DSS 的结构体系及开发[J]. 莱阳农学院学报,2000,17(1):67-70.
- [23] 郭朝珍,杨俊杰. 基于数据仓库技术和 OLAP 的决策支持系统[J]. 系统建设,2000(1):11-13.
- [24] 林明. 协同:技术概念不断延扩技术手段不断丰富[EB/OL]. http://industry.ccidnet.com/art/10/20060411/503117_1.html.
- [25] 搜索引擎原理[EB/OL]. <http://www.3lian.com/zl/2004/10-26/235555.html>.
- [26] Web2.0 教程[EB/OL]. <http://soft.yesky.com/327/2055327.shtml>.
- [27] IBM 门户技术与实践[EB/OL]. <http://www.ibm.com/software/cn/websphere/portalwebcast/>.