

●胡昌平 汪会玲

信息资源的异构整合及其对平台建设的要求*

摘要 要以满足用户个性化信息需求为中心开展信息资源整合平台建设,要变纵向整合为横向纵合。整合后的信息资源平台,主要由用户平台和资源整合平台组成。图3。参考文献5。

关键词 信息资源 异构整合 整合平台 个性化

分类号 G253

ABSTRACT To meet use's demand for personalized information, we should develop information integration platforms, which should consist of user platforms and resource integration platforms. 3 figs. 5 refs.

KEY WORDS Information resource. Heterogeneous information resource. Integration platform. Personalization.

CLASS NUMBER G253

信息资源的异构整合为解决多平台、多结构信息资源的集成问题提供了一条解决途径,专家学者们也从各个方面对它进行研究,并取得了丰富成果,然而对信息资源整合中用户的个性化信息需求研究却相当匮乏。从用户的个性化信息需求出发,构建一致的、能广泛共享的信息资源整合平台,进而在此基础上实现系统内外各种相关资源的整合,对提升信息资源整合整体效果有很好的促进作用。

的问题逐步从如何查找信息转变为如何从信息海洋中筛选出自己所需要的信息,简单的信息资源共享已经不能满足用户需求。网络信息用户希望互联网能够给他们提供一种经过筛选、整合、优化之后的个性化信息服务,可以预见,个性化信息服务将成为信息服务发展的新方向。在个性化信息服务中,为满足用户对信息的“精、准”要求,需要对信息资源精加工。信息资源整合是信息资源精加工的重要工具,为个性化信息服务提供了很好的资源保障^[1]。

1 信息资源集成化与服务个性化发展

随着网络信息资源爆炸式地增长,用户面临

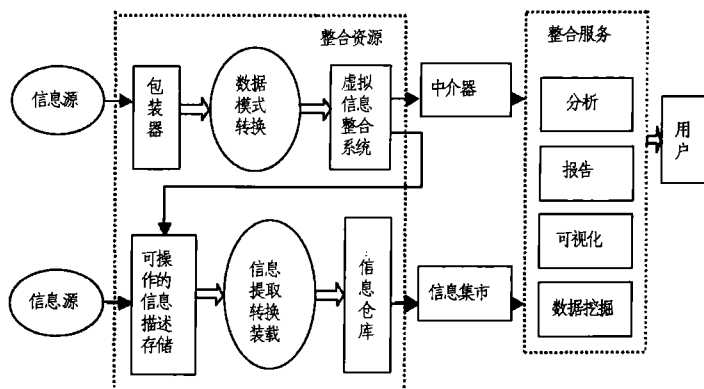


图1 个性化信息服务中的信息资源整合

* 本文系国家自然科学基金项目“面向用户的信息整合与服务研究”(编号70373049)的研究成果之一。

由图1可知,个性化信息服务中的信息资源整合主要体现在两个方面。其一是整合资源,即通过一个门户入口提供知识导航,利用信息技术,如元数据、数据仓库、知识挖掘等技术,对现有各种异质的、不一致的信息资源进行重组整合,从中提取能给用户解决问题的知识,并以用户能够接受的方式发送给用户;其二是整合服务,即利用计算机、网络技术为用户提供全方位、一站式的个性化服务,如对数据仓库中的信息进行关联分析、数据挖掘并以可视化的方式提供给用户。

2 信息资源的异构整合平台设计要求

2.1 以用户的个性化信息需求为中心开展信息资源整合平台建设

为满足用户的个性化信息需求,在信息资源整合平台建设中应特别注重以下两点。

首先,变面向信息的信息资源建设为面向用户的信息资源建设。受传统信息管理和服务的的影响,我国信息资源建设的目标主要集中在对信息本身的收集、整理、加工和存储方面,研究重点也主要集中在信息组织的机理与技术研究。从中国期刊网的检索数据可以明显看出这种趋向。我们以篇名检索“信息资源”可以检索出5292篇(时间从1994年1月~2004年7月)。而我们以篇名作第二次检索“用户”则只有22篇,以关键词进行第二次检索也才198篇,以中文摘要进行二次检索“用户”也只有202篇。由此可见,在信息资源建设的研究中强调用户需求研究的非常少,最多也不过3.8%。而在IEEE中,以篇名检索“Information Resource”可得35篇;以摘要进行二次检索“User”的有11篇,占总数的31%。可见中外对“用户”研究的差距很大。而面向用户的信息资源建设以及在此基础上开展的个性化信息服务是未来信息资源建设和信息服务的发展方向^[1]。基于个性化信息服务的信息资源整合平台建设也应该从用户的需求着手,着重于面向用户的信息资源建设。

其次,要从用户角色分析入手细分团体用户的

需求特征。但建立用户信息库难度非常大,因为很难得到用户的个人信息,如职业、职称、爱好,甚至姓名与年龄,用户都不大愿意提供。由于涉及个人隐私,也不便收集。而在个性化信息服务中,系统获取的用户信息越多,就越能根据用户需求提供高相关度的信息。比较有效的办法就是细分用户群,将用户纳入用户群组,为不同的群组成员提供不同的信息服务内容及其方式。这样,一来可满足用户的相对个性化的需求,二来有利于系统组织和优化信息资源。可以从用户的角色入手来划分用户群体,例如:教师、学生、博士、硕士、本科生。现实生活中,用户往往扮演多重角色,譬如一个博士研究生可能是一位老师,也可能是一位家长,还可能是一位音乐爱好者。为了能更好地将用户归于某一适当的用户群组,可以按照影响用户信息需求因素的重要性来进行角色排序,如用户职业、用户的职责、用户所受的教育、用户兴趣爱好、用户个人的信息素养等几个影响用户个体信息需求的主要因素。可以从职业、年龄、学历、爱好等几个方面逐层地对用户群进行划分。

2.2 变纵向整合为横向整合

在信息资源整合的研究中,对信息内容整合的研究,对企业机构的整合,以及服务整合等进行单独研究的比较多。大多突出信息资源的合作开发和整合,而没有从信息用户的全方位需求出发,将系统资源、机构、人力资源、服务环境等整合起来综合考虑,这样从某种程度上又造成了资源建设与服务以及用户需求的脱节。为充分发挥信息资源整合的最大效益,应该打破这种整合格局。从用户服务组织的机理出发,对机构、服务、用户和资源进行重组,使用户和系统的交流更顺畅,使系统资源得到更充分利用。我们并不是否定以前的信息整合、机构整合、服务整合的研究成果,而是在这些整合基础上,将以上几方面的整合串起来,使之成为信息资源平台建设系统的有机组成部分,真正形成用户、服务、资源一体化建设(如图2)。

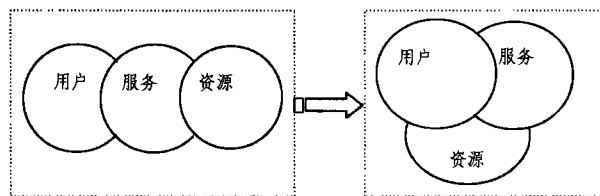


图2 用户、服务、资源的链式模型与交互模型

在目前的串行模型中,用户的信息需求只有通过服务组织传递给资源系统,系统资源对用户的需求进行处理后再交给用户,这种通过中间环节“服务”进行的交互在服务效率与效果上都有所降低^[2]。如果将用户信息直接纳入资源建设的一部分,利用数据挖掘技术对用户信息进行分析,就可以直接将用户的一些需求处理结果传递给用户。例如在期刊服务中,系统可以自动分析用户检索时失败次数及原因。假如一些用户总是提交最近3个月期刊文献的检索要求而检索失败,系统资源建设部可考虑是否应该自建一个现刊目次查询库以满足用户的信息需求;若发现检索时间过长的用户往往频繁使用不规则的检索词,系统资源建设部则可考虑通知服务部是否可就各专业学科进行单独的检索培训。这种用户、服务、资源互为一体的信息资源平台建设系统,将会大大提高服务效率和效果。

3 基于个性化服务的信息资源整合平台实现

3.1 基于个性化服务的信息资源整合平台构建

任何网络信息提供者要为用户提供服务,必须涉及网络平台、硬软件平台、数据资源平台、交换平台、业务平台与服务平台^[3]。为满足用户的个性化信息需求,不仅要对这几种服务平台进行纵向整合,更重要的是从用户的需求及其兴趣、偏好出发,对这些平台进行横向整合,变面向信息资源的资源建设为面向用户、面向服务的信息资源建设。这就要求重组现有的网络资源、信息资源、人力资源、服务业务等。整合后的信息资源平台主要由两部分组成:用户平台和资源整合平台(如图3)。

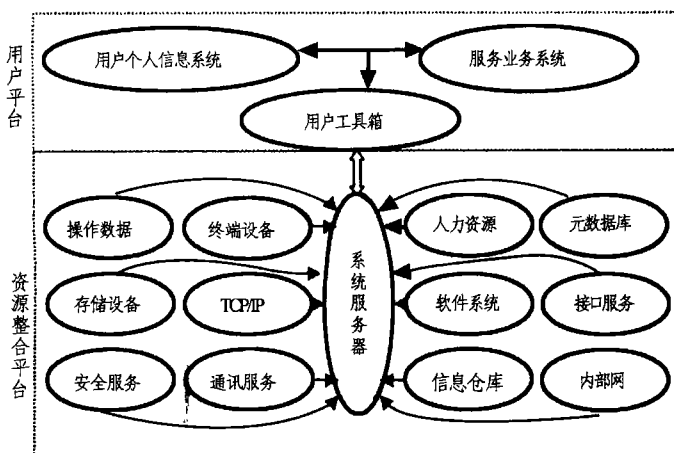


图3 基于个性化服务的整合平台结构

其中,用户模块主要包括用户信息与服务业务两大块,应从用户需求的角度对二者进行整合并通过用户工具箱与整合平台系统资源进行交互。用户个人信息系统包括用户基本信息、用户需求信息、用户定制信息、用户管理信息。用户基本信息主要是指用户ID、用户名、密码、真实姓名、电子邮件、电话、邮编、家庭地址、工作地址等用户最基本的信息。用户需求信息包括用户资源定制、用户服务方式的定制,以及用户服务界面的定制。用户管理信息包括用户元数据管理、用户远程登录管理、用户访问时间与次数管理等。服务业务系统包括所有的传统与现代信息服务

的整合。资源整合平台则打破传统的条条框框管理体制,对已有的系统资源包括信息资源、系统软硬件资源、人力资源等进行重组整合^[4]。例如:用户提供有关元数据方面的信息需求,系统调用用户信息库之后,将需求请求一次性地传递给各设备,一次性完成请求处理并将处理结果提供给用户。要特别注意人力资源和软硬件设备资源的整合,做到系统各种资源的优化配置与使用。

3.2 基于个性化信息服务的信息资源整合平台建设的实现

该平台模型可采用B/S模式实现,在客户端实

现用户个人信息系统设计、服务业务整合等,在服务器端实现基于信息资源整合系统的系统资源整合,即在信息资源整合系统的基础上,整合组织内已有的系统资源,如 CRM 系统资源、管理信息系统资源、供应链系统资源、财务信息系统资源等,实现组织内信息、人力、资金、物资、客户等资源的整合管理^[5]。

在客户端用户平台的设计中,用户信息的收集和服务业务整合均在客户端。可以充分利用用户个人信息定制实现个性化的数据展现,将资源整合平台返回的 XML 查询结果高效地转化为用户喜欢的样式,如饼图、表格、云图、柱状图等。客户端的用户信息资源可以采用基于 XML 的 RDF (resource definition framework) 来组织,并利用支持 XML 的数据库系统来存储用户信息资源,这样不仅利用了 XML 的优点,也保持了系统的性能。在用户信息采集,Java 语言提供了对 XML 强有力的支持,如 JAXP, JDOM, Java XML 解析器,利用它们提供的服务很容易将用户信息封装成 XML 形式或者对外部传来的 XML 数据进行解析,实现用户个人信息系统与服务业务系统之间的整合。

在服务器端,该集成平台框架基于 J2EE 标准,采用 Java 编程语言实现,利用它与平台无关、复用性强等特点,实现包括 XML, dbf, Access, SQL Server, Sybase, Oracle 等在内的各种系统资源的整合。采用 XML 作为集成模型描述语言,隐藏了每个数据源的局部特性,而且 XML Schema 模型的充足性简化了各资源模型之间的映射。使用功能强大的 XML 统一查询语言 Xquery,使基于标准接口用统一的方式查询 XML 全局和局部视图成为可能,以此来增强平台的整体性能。

本解决方案具有以下优势。

安装过程更加简便,网络结构更加灵活,B/S 模式对浏览器界面显示和交互功能的充分利用简化了客户端,使得客户端可以是网络上的任何操作平台。

扩展性、可维护性好。所有的数据管理均集中在服务器端,使开发、维护过程中的大部分工作可以集中在服务器端,无需考虑客户端的系统状况,仅在调用和调试时涉及客户端。用户信息系统与服务系统的整合,使得系统能够方便地增加或调整网络服务类型以适应用户新的信息需求,并且使用 Java 技术,使系统具有良好的可移植性。

系统适应性强。该整合平台既能够适应 Oracle, Sybase, DB2, SQL Server, Mysql, Access, dbf 等多种结构化数据源的整合,又能够适应 XML 等半结构化数据源的整合。

用户界面友好。利用与客户端环境结合的优势,可以从更多的方面观察并获取与用户相关的特征信息,并运用这些用户特征信息参与信息检索,帮助搜索引擎获得与用户偏好更贴切的个性化信息。

参考文献

- 1 杨元庆. 构建新型互联网应用模式. 中国计算机用户, 2002(19)
- 2 Hendrik Edelman and Robert P. Holley. Marketing to Libraries for the New Millenium. Maryland: Scarecrow Press, Inc, 2002
- 3 Yousuke Watanabe. Integration of Multiple Dissemination-Based Information Source Using Source Data Arrival Properties. Web Information Systems Engineering, 2001(5)
- 4 Stella Pilling and Stephanie Kenna. Cooperation in Action—Collaborative Initiatives in the World of Information. London: Facet Publishing, 2002
- 5 王瑞等. 基于 B/S 模式的远程数据采集系统的研究. 计算机应用, 2003, 23(4)

胡昌平 武汉大学信息资源研究中心教授, 博士生导师。通信地址: 武汉。邮编 430072。

汪会玲 武汉大学信息管理学院 2003 级博士研究生。通信地址同上。(来稿时间: 2005-08-22)

欢迎订阅和零购《中国图书馆学报》

《中国图书馆学报》是文化部主管, 中国图书馆学会和国家图书馆主办的国家级图书情报学专业期刊, 被评为中国优秀图书馆学期刊、全国中文核心期刊、中国期刊方阵期刊和国家期刊奖百种重点期刊。

全国各地邮局均可订阅, 既可全年也可被年订阅, 国内代号 2-408。中国国际图书贸易总公司负责国外发行, 国外代号 Q184。2006 年每期定价 13 元, 全年 78 元。本刊编辑部零售。

也可在本刊编辑部订阅或购买单册。在编辑部直接订购的个人订户, 每期优惠价 10 元(含寄费)。编辑部地址: 北京中关村南大街 33 号。邮编 100081。电话 88545141。订阅方式可以直接汇款, 在汇款单上写明订阅者名称、地址、邮编、份数。也可银行转账。开户名称: 中国图书馆学会。开户银行: 北京银行魏公村支行。账号: 01090303200120105049050。