

●郭海明¹, 邓灵斌²

(1. 潍坊学院 图书馆, 山东 潍坊 261061; 2. 武汉大学 信息管理学院, 湖北 武汉 430072)

面向用户的数字图书馆信息服务系统构建*

摘要: 面向用户的数字图书馆信息服务系统是一个以网络化数字化知识资源为基础, 以用户需求为中心, 以先进的信息技术为手段的资源、技术、服务与用户利用一体化的数字资源平台。作者对其系统总体设计、具体模型、关键技术与问题进行了探讨。

关键词: 数字图书馆; 信息服务系统; 系统构建

Abstract: The user-oriented digital library information service system is a digital resources platform based on networked and digitalized knowledge resources, centered on users' need and integrating resources, technology, service and users' utilization with the advanced information technique as the means. This article discusses the overall design, specific model, key technologies and existing problems of the system.

Keywords: digital library; information service system; system construction

信息服务系统的发展经历了面向资源型、面向交流型、面向用户型的演化过程。面向资源型是指信息服务系统的设计与工作中心是以信息资源的加工处理为核心; 面向交流型是指信息服务系统的设计与实现以信息的快速有效传递为核心; 面向用户型是指信息服务系统的设计与实现以用户利用信息的需求为核心。未来的图书馆信息服务系统特别是数字图书馆信息服务系统将是一种面向用户型的信息系统。面向用户的数字图书馆的服务系统是一个以网络化数字化知识资源为基础, 以用户需求为中心, 以先进的信息技术为手段的资源、技术、服务与用户利用一体化的数字资源平台。

1 系统构建的总体目标

面向用户的数字图书馆信息服务系统构建的目标是建设一个既符合网络传输特性又符合人们接受数字信息习惯的服务系统, 合理配置与协调信息服务系统中的各组成要素及其与所处的环境的关系, 使它们发挥最大的系统功效, 为信息用户提供所需的信息与服务, 满足信息用户学习与解决现实问题所需的知识和信息需求。面对分布、异构和自主的数字资源与服务, 数字图书馆信息服务系统需要对这些资源与服务进行集成, 通过互操作、整合处理以及集成管理, 提供面向用户的逻辑上统一或相互有连接的集成数字信息服务, 支持用户在整个网络信息环境中有效

搜寻、获取和利用数字信息^[1-2]。

1) 集成性。面对分布、异构和自主的数字资源与服务, 数字图书馆信息服务系统要能够实现对这些资源与服务的集成, 通过互操作、整合处理、集成管理提供面向用户的逻辑上统一或相互有连接的信息集成服务, 支持用户在整个网络信息环境中有效搜寻、获取和利用信息与知识。

2) 互操作性。系统要能够实现面向用户屏蔽分布的、异构的各个数字图书馆之间的差别, 提供一致化的检索和服务。

3) 开放性。系统应该是一个开放的体系, 对内各模块(包括仓储和服务)可以随进随出, 对外能和其他的系统交互, 取长补短; 支持丰富多样的资源和功能形式, 能容纳各种各样的信息资源体系和服务系统; 支持分布的各个系统的自主建设与发展, 支持专门的甚至本地化的元数据格式和系统协议。

4) 可扩展性。系统的体系结构应该是灵活的和可扩展的, 能随时接纳新的仓储和服务, 而不需要大的改动, 保证整个服务系统的可伸缩性, 能容纳甚至动态组合任意数量和类型的信息资源或服务系统。

5) 低成本性。必须保证整个服务系统的低成本和进入该系统的低成本; 保证在系统中增加新资源或服务系统的方便性和低成本; 保证在系统中使用任何一个资源或服务系统的方便性和低成本。

6) 系统友好性、可嵌入性。面向用户的数字图书馆信息服务系统要有友好的系统界面, 而且服务系统要能与

* 本文为潍坊学院哲学社会科学科研项目“基于用户的数字图书馆信息服务研究”成果论文。

用户的工作系统与环境友好对接与兼容,数字图书馆信息服务系统可以直接嵌入到用户的计算机工作环境中去,成为用户工作系统的一个组成部分。

2 系统总体结构与工作模式

2.1 体系结构

数字图书馆信息服务系统的体系结构仍在不断完善和发展之中。面向用户的数字图书馆信息服务系统也应该是一种三角形的三层客户/服务器模式的“浏览器—Web 服务器—数据库服务器”结构,如图 1 所示。数据库服务器、Web 服务器和浏览器三部分分别对应于数据库层、中间层(业务逻辑层)和客户层(表示层)。在三层结构中,每一层支持应用程序的一个独立部分。Web 服务器主要接收读者客户端的查询请求,进行数据处理和处理结果的发送,管理 HTML 构成的信息空间,提供对数据库的存取接口;数据库服务器主要负责管理数字化馆藏,通过接收 Web 服务器请求,对数据进行处理,然后把处理结果传送给 Web 服务器;读者客户端通过各种网络实现与 Web 服务器的连接,通过浏览器访问 Web 服务器提供的各种功能和丰富的数字化馆藏^[3,4]。

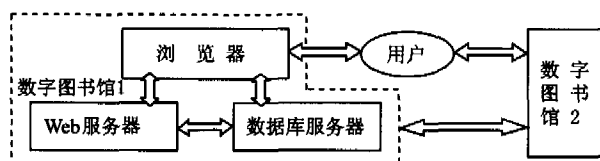


图1 数字图书馆体系结构图

2.2 系统工作模式

面向用户的数字图书馆是一种基于用户信息活动的数字信息资源管理与服务系统。它是以一个无所不在的分布式的网络和计算机为技术支持和应用基础的,基于信息资源的服务形态,围绕用户信息活动和用户信息系统来组织、集成、嵌入数字信息空间的信息资源和信息服务,针对不同需求的用户群体提供智能的、主动的、个性化的终极信息服务,支持用户灵活自助自主地处理信息、提炼知识、协作交流和解决现实问题的一种数字信息资源管理与服务系统^[5]。这一数字图书馆的信息服务模式应该是一种“用户中心”服务模式。具体地说,是一种基于用户信息活动面向问题服务,主动与利用自助相结合的资源分布、服务集成的“用户中心”服务模式^[6]。它生存于一个聚合化数字信息空间环境里,基于并全程跟踪用户信息活动,强调服务者与用户主体双方的主观能动性,把数字图书馆信息服务系统嵌入到用户工作与信息利用环境中,

克服了基于信息资源的服务模式中将信息系统与用户信息利用过程相对隔绝的局限,有助于用户在信息利用过程中对数字信息对象的灵活处理、知识提炼和协作交流。围绕用户信息活动、用户需要解决的现实问题和用户信息系统来组织、集成、嵌入数字信息资源和信息服务,从而更直接、深入、有效地支持用户自助检索、处理、利用信息来解决现实问题的全过程,最终达到用户满意,从而增进用户利益超越用户期望^[7]。见图 2。

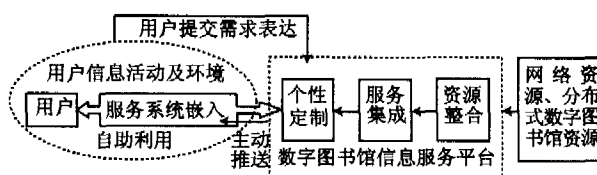


图2 数字图书馆信息服务未来模式图

3 系统逻辑模型总体构建

面向用户的数字图书馆信息服务系统的逻辑模型包括 4 个层次:服务层、功能层、数字对象层、平台层。①用户服务层。处于数字图书馆信息服务系统的最高层,是各类数字图书馆信息用户的高层应用表现。②功能层。处于服务层与数字对象层之间,是数字图书馆信息服务系统各种管理与服务功能的具体实现,是该服务系统的核心。③数字对象层。是数字图书馆信息服务系统的信息资源的储存中心。④平台层。处于数字图书馆信息服务系统的最低层,是系统的技术与管理支撑^[8]。见图 3。

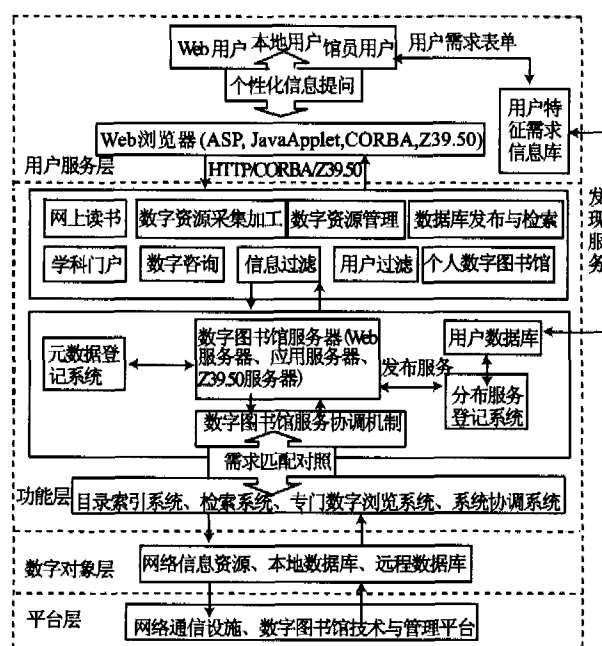


图3 数字图书馆信息服务系统逻辑模型

4 系统逻辑模型实现的关键技术与问题

4.1 元数据问题

元数据 (Metadata) 被称为关于数据的数据 (Data about other Data), 它是专门用来描述数据的特征和属性的、提供某种资源的有关信息的结构数据 (Structured Data), 也就是说, 元数据是促进数字图书馆信息资源的组织和发现的数据。元数据能改进数字信息资源的检索能力 (特别是检索的精确性), 以及信息资源的控制和管理问题。上述逻辑模型中的元数据登记系统是面向用户的数字图书馆信息服务系统中信息的描述方法, 它决定了系统的特征、运行模式和系统运行的总体性能。因此, 元数据体系是数字图书馆资源管理系统和网络导航系统的管理和控制层, 在数据挖掘、信息检索和信息组织方面都有着重要的作用。我们一定要设计好它的元数据体系。

4.2 互操作性问题

数字图书馆的互操作是指共享数字图书馆的文档、查询和服务的能力, 它要求来自技术 (消息交换的协议和格式等)、内容 (数据、元数据及其语义等) 和组织 (数字图书馆的基本访问原则、支付方式和认证等) 三方面的合作。要有效利用分布式数字信息环境中的各类信息资源和信息服务, 必须保证分布的数字信息系统间的互操作性 (Interoperability), 即这些分布系统之间能无缝地交换、共享信息资源和信息服务, 并能在不损害各个分布系统自主性的同时构成一个 (往往是虚拟的) 集成系统逻辑机制。面向用户的数字图书馆信息服务系统的互操作应该满足: ①支持丰富多样的资源和功能形式, 能容纳各种各样的信息资源体系和服务系统。②支持分布的各个系统的自主性, 能持续支持各个系统的自主建设与发展, 能有效保证对知识产权资源使用的本地控制, 能支持专门的本地客户端机制和服务, 能支持专门的甚至本地化的元数据格式和系统协议。③保证整个分布式数字信息机制的低成本和进入该机制的低成本。④保证在分布式数字信息机制中增加新的资源或服务系统的方便性和低成本。⑤保证在分布式数字信息机制中使用任何一个资源或服务系统的方便性和低成本。⑥提供整个机制的可伸缩性, 能容纳甚至动态组合任意数量和类型的资源或服务系统^[9]。

4.3 用户资源管理问题

数字图书馆信息服务系统若要真正实现“以用户为中心”, 必须采用有效的方式来收集用户的相关信息, 并对用户信息进行分析, 以明确用户的信息需求, 建立用户信息需求模型。然后通过智能代理, 主动从网络信息资源或数据库中搜索信息。一旦出现符合用户需求条件的最新信息, 就将这些信息进行筛选、分类、排序, 利用信息推

送技术及时主动地推送给用户, 从而实现从传统的“人找信息”到“信息找人”的突破性变革, 使数字图书馆信息服务系统成为用户获取各种网络资源与服务的个性化门户。收集和管理用户资源的主要途径有: ①通过个性化设置获取用户兴趣资源。②通过服务器日志分析用户需求资源。③根据用户检索结果, 利用数据挖掘的原理, 提取网页的主题特征, 并将提取出来的特征项作为用户兴趣的组成要素, 丰富用户兴趣模型。④通过对用户查询行为进行记录而实现用户资源的搜集。

4.4 信息过滤技术

信息过滤 (Information Filtering) 是根据用户的信息需求, 运用一定的标准和工具, 从大量的动态网络信息流中选取相关的信息或删除不相关信息的过程。通过对用户检索和其他交互过程的分析, 建立用户个性化的模型描述, 利用此模型和多媒体信息空间的匹配, 根据一定的要求设置过滤条件, 在过程中一旦触发条件则将有关的信息拒之门外, 而其他信息可以进入, 实现个性化的知识服务。它在为用户提供所需要信息的同时, 着重剔除与用户不相关的信息, 从而提高用户获取信息的效率; 它根据用户信息需求的变化提供稳定的信息服务, 能够节约用户获取信息的时间, 从而极大地减轻用户的认知负担, 起到减压阀的作用。信息过滤技术基本分为两类: 一种是基于内容的过滤 (Content-based Filtering); 另一种是合作过滤 (Collaborative Filtering)。综合基于内容和合作过滤两种方式的优点, 面向用户的数字图书馆信息服务系统应采用基于混合模式的信息过滤 (Hybrid Filtering) 模型。它建立面向个人的用户模板和面向合作的公共模板, 抽取信息特征, 作为可能的特征项, 便于用户动态地修改模板; 利用其他用户对文档的评价以及用户模板与文档的相似度来预测用户的接受程度, 另外还考虑到推荐者的权威性和与用户兴趣的一致性^[10]。其处理流程如图4所示。

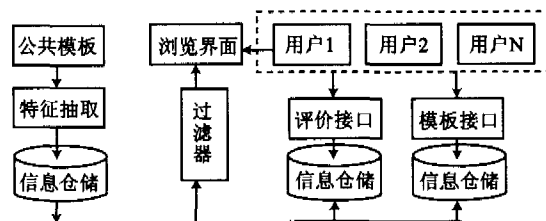


图4 基于混合模式的信息过滤模式

4.5 学科信息门户技术

学科信息门户是将特定学科领域的信息资源、工具与服务集成到一个整体服务系统中, 为用户提供一个方便的信息检索和服务入口。面向用户的数字图书馆信息服务系

统必须致力于学科门户的建设。具体有以下 5 种形式：①主题信息门户。以网络学科信息导航为主的，提供权威、可靠、规范和可持续的网络信息资源选择、描述和检索。②专业信息门户。以专业机构或图书情报服务系统为基础，根据专业机构性质或其信息服务要求，将各类资源（包括网络资源、数据库、文件系统、知识库、指南手册等）组合在统一门户下向用户提供服务。③跨学科信息门户。基于和支持多个学科信息门户之间的整合检索。④分布信息门户。基于信息门户体系，将多个分布门户整合成一个集成门户体系，让用户通过信息门户体系方便地搜寻、调用和利用各种不同的信息资源和服务。⑤开放信息门户。不但支持基于学科信息门户的资源与服务集成，还进一步支持按照用户个性化需要定制信息门户，根据逻辑业务流程整合多个信息服务环节，支持多个信息门户之间的开放集成与定制^[11]。见图 5。

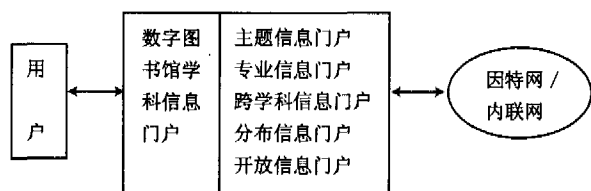


图 5 学科门户服务工作流程图

4.6 信息智能代理技术

智能代理技术是一种能够完成委托任务，模仿人的行为执行一定的任务，不需要或很少需要用户的干预和指导的智能计算机系统。面向用户的数字图书馆信息服务系统通过智能代理跟踪用户在信息空间中的活动，自动捕捉用户的兴趣爱好，主动搜索可能引起用户兴趣的信息并提供给用户。利用信息智能代理技术是实现面向用户的数字图书馆信息服务系统个性化信息服务的有效方法^[12]。其主要功能有：①个性化的信息管理代理库；②信息自动通知；③浏览导航；④智能搜索；⑤动态个性化页面。

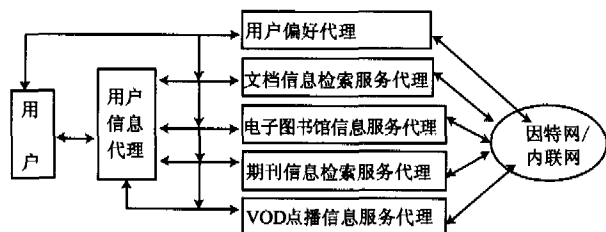


图 6 智能代理服务工作流程图

一种基于客户端信息智能代理模型的工作流程描述如图 6 所示。“用户信息代理”（Agent Server）模块列出数

字信息服务系统可以提供的功能选项，用户从中选择自己所需要的信息服务功能，并且选择对各种信息服务所期望的显示方式。用户的功能需求信息会被传递到代理服务器（Agent Server）来构造个性库（Individuality Database）；而显示方式则存储于本地“用户偏好”（User Preferences）中。个体请求的信息及 Agent Server 主动提交的信息反馈到各种服务代理（Agent，如文档信息检索服务 Agent）时，就会结合 User Preferences 中事先预定义的显示要求提交给用户。□

参考文献

- [1] 张晓林. 分布式数字图书馆机制 [J]. 情报学报, 2002 (1): 63-70
- [2] Paepcke A, et al. Interoperability for digital libraries world wide [J]. Communications of the ACM, 1998, 41 (4): 33-43
- [3] 盛小平. 国内外数字图书馆发展的比较研究 [J]. 中国图书馆学报, 2001 (6): 39-44
- [4] 耿骞. 基于元数据的数字图书馆信息系统的研究 [D]. 北京: 北京师范大学, 2001
- [5] 张晓林. 数字图书馆机制的范式演变及其挑战 [J]. 中国图书馆学报, 2001 (6): 3-8, 17
- [6] 郭海明, 邓灵斌. 数字图书馆信息服务模式研究 [J]. 中国图书馆学报, 2005 (2): 47-49, 53
- [7] William Y A, Christop B. An architecture for information in digital libraries [J]. D-Lib Magazine, 1997 (February)
- [8] 郭海明, 刘昆雄. 数字图书馆信息服务系统构建探讨 [J]. 图书馆学、信息科学、资料工作, 2004 (5): 141-145
- [9] 赵四友. 数字图书馆互操作性问题初探 [J]. 现代图书情报技术, 2002 (2): 8-10, 24
- [10] 刘柏嵩. 过滤研究 [J]. 现代图书情报技术, 2003 (6): 23-26
- [11] 张晓林. 分布式学科信息门户中网络信息导航系统的规范建设 [J]. 大学图书馆学报, 2002 (5): 28-33, 43
- [12] 庄鹏, 等. 代理模式实现数字图书馆个性化信息服务模型 [J]. 情报学报, 2004 (2): 25-30

作者简介: 郭海明, 男, 1973 年生, 硕士, 馆员。发表论文 20 余篇。

邓灵斌, 男, 1972 年生, 博士生。发表论文 20 余篇。

收稿日期: 2006-04-09