●张银犬　朱庆华

网格环境下个人数字图书馆信息检索策略*

摘　要　P2P 技术和 Z39.50 协议在个人数字图书馆信息检索策略模型中有机结合,有助于充分实现分布式个人数字图书馆的检索、交互功能。构建 P2P 个人数字图书馆网络,实现个人数字图书馆信息检索,需要利用中间件技术将 P2P 技术、Z39.50 协议(包括 SRW/SRU)和个人数字图书馆软件捆绑在一起。图 2。参考文献 13。

关键词　个人数字图书馆　信息检索　Z39.50 协议　Peer to Peer
分类号　G250.76

ABSTRACT　The integration of P2P technology and Z39.50 protocol in the model of information retrieval strategies for personal digital libraries can help to realize the retrieval and interactivity of distributed personal digital libraries. To construct P2P personal digital library networks and to realize personal digital library information retrieval, we should use midware technology to bind P2P technology, Z39.50 protocol (including SRW/SRU) and personal digital library software systems together. 2 figs. 13 refs.

KEY WORDS　Personal digital library.　Information retrieval.　Z39.50 protocol.　Peer to peer.
CLASS NUMBER　G250.76

从理论角度分析,个人数字图书馆能够成功实现对存贮在个人计算机中的信息资源(知识)进行有效组织与管理,但如何成功地实现对网格环境下分布的个人数字图书馆的信息检索,并实现彼此之间的共享,还是一个研究空白与难点。本文针对个人数字图书馆信息检索,提出一个检索策略模型,试着将个人数字图书馆架构在 P2P(Peer to Peer)网络之上,创建一个分布式的信息检索环境,并采用 Z39.50/SRW/SRU 信息检索协议实现个人数字图书馆信息检索,探讨分布式个人数字图书馆信息检索机制,为网格环境下个人数字图书馆信息检索技术的开发提供理论基础。

1　个人数字图书馆的概念、本质及功能

个人数字图书馆是数字图书馆发展中的新鲜事物,反映了网络环境中信息用户的个性化需求,是网络信息资源管理的客观要求,也是知识管理发展的必然。作为一个新生事物,国内外研究者对个人数字图书馆概念的理解并不一致。

陈光祚教授的定义被大多数个人数字图书馆研究者引用,即"指个人为了读书治学的目的,在自己的计算机上采用免费或基本免费的全文数据库软件,将有关的网上信息和自创的数字化信息资源进行采集、存储、使之成为有组织的信息集合,以供自己个人有效利用的数字图书馆"。他将个人数字图书馆比做是"e"时代的私人藏书楼[1],也被国内研究者所接受,充分反映了用户的个性化需求。

个人数字图书馆的内部结构与外部联系必然发生变化。本文认为:个人数字图书馆是指采用现代信息技术对个性化需求信息进行管理,是实现个人知识管理的一种工具,是公共数字图书馆发展与服务延伸的个性化数字图书馆。个人数字图书馆在功能上满足了信息用户的个性化需求,实现了个人信息管理、知识管理。个人数字图书馆离不开网络,必须充分实现与公共数字图书馆和其他个人数字图书馆之间的协同,是动态交互的知识管理平台。它的本质应当是"个性化"的数字图书馆,而非"个人的"数字图书馆。在倡导开放存取的时代,个人数字图书馆功能应当是为社会服务的、开放的、交互的、共享的。

2　P2P 技术及其在分布式个人数字图书馆信息检索中的应用

2.1　分布式个人数字图书馆特征和技术需求

个人数字图书馆创建模式多样,但国外主流是将它构建在个人计算机上。个人计算机上构建个人数字

* 本文为南通大学人文社科项目(03040485)、江苏省教改课题(J0541)的成果之一。

图书馆软件很多,如:超级文档管理器、网海拾贝、良友收藏家、ELIB 电子图书馆、TRS 个人信息中心、MYBASE、ADKSAM4 等[2],它们采用元数据方法对存贮在个人数字图书馆中的信息资源组织与管理,为成功实现个人数字图书馆中信息资源的检索做了必要准备。

分布式的个人数字图书馆建立在个人计算机上,具有极大的分散性,无法通过传统的 C/S 模式来实现信息资源的检索,客观上需要构建一种分布式的个人数字图书馆网络,同时又必须具备去中心化、自由化等特征。

个人数字图书馆必须具有共享、互操作功能,客观上要求能对分布式的个人数字图书馆信息资源进行共享检索,发挥个人数字图书馆的功能,现有个人数字图书馆软件不能实现。因此还必须寻求其他的技术支撑。P2P 网络具有以上特征,在某种程度上使个人数字图书馆网的构建与 P2P 技术相结合成为可能。

2.2　P2P 基本概念及关键技术

传统因特网实现了计算机硬件的连通,Web 实现了网页的连通。而网格试图实现互联网上所有资源的全面连通,包括计算资源、存储资源、通信资源、软件资源、信息资源、知识资源等,它强调全面地共享资源、全面地应用服务,目前采用的是 P2P 计算体系结构。

P2P 一般称作对等网,它是一种网格,也是一种技术。IBM 认为 P2P 系统是由若干互联协作的计算机构成,系统依存于边缘化(非中央式服务器)设备的主动协作,每个成员直接从其他成员而不是从服务器的参与中受益;系统中成员同时扮演服务器与客户端的角色;系统应用的用户能够意识到彼此的存在,构成一个虚拟或实际的群体。P2P 网络是互联网整体架构的基础,在通信过程中,所有的设备都是平等的一端[3]。P2P 技术改变了"内容"所在的位置,使其从"中心"走向"边缘",也就是说内容不再存于主要的服务器上,而是存在所有用户的个人计算机上。P2P 使得个人计算机不再是被动的客户端,而成为具有服务器和客户端双重特征的设备。

以网络中有无服务器为依据,可将对等网分成混合式 P2P 网络和纯分散式 P2P 网络两大拓扑结构。典型系统有 Napster, Gnutella, FreeNet, Chord-based System, BitTorrent 等。P2P 系统包括 P2P 平台层和应用层。P2P 平台层包含支撑 P2P 应用所需的基础组件,如发现机制、通信、安全、资源集成等组件。其应用层利用 P2P 平台提供的功能,向用户提供专门的服务。Peer 通信和定位、平台的安全和平台的性能优化是 P2P 系统成功与否的关键技术。有学者从网络的拓扑结构和 Peer 节点的角色划分、资源的标识、Peer 的定位方式、P2P 网络中节点的登录、退出和节点故障、防火墙和 NAT 的穿越、P2P 平台的安全机制、P2P 平台的性能改善技术等方面进行详细论述[4]。

2.3　利用 P2P 技术构建分布式个人数字图书馆网络

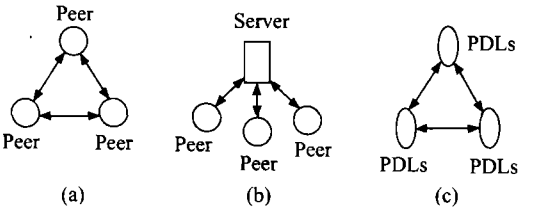

图 1　有关的拓扑结构

P2P 对等网络存在纯分散式与混合式两种网络拓扑结构,分别如图 1(a)和(b)所示。个人数字图书馆网的构建采取 P2P 网络的哪种模式取决于个人数字图书馆发展的具体实际情况。高度分散性特征,使得个人数字图书馆的发展无论在主观上,还是客观上,均很难产生一个统一集中的 Server。因此,选择图(c)——纯分散式拓扑结构的个人数字图书馆对等网模式,将与图 1(a)中纯分散的 P2P 网络拓扑结构相一致,其中 PDLs(Personal Digital Libraries)相当于 Peer,不同的 PDLs 通过 P2P 网络实现信息资源共享、检索等功能。

2.4　P2P 技术在分布式个人数字图书馆信息检索中的功能

P2P 技术在实现对等点的检索方面有强大功能。如:集中的搜索机制、分散的搜索机制和分散且结构化的搜索机制,移动 Agent 技术[5];广度优先搜索、随机广度优先搜索、定向广度优先搜索[6];基于 XML 的 DHT 索引检索机制[7];基于推荐策略的搜索算法 RPSA[8];基于语义路由的 P2P 检索系统体系结构等等[9]。

但在个人数字图书馆对等网中,P2P 技术并不用来实现具体的信息检索,而是用来互联、定位、发现个人计算机中建立的个人数字图书馆,为个人数字图书馆信息检索做必要准备。

P2P 技术在个人数字图书馆对等网中的功能主要体现在:(1)建立个人数字图书馆之间的通信机制,实现个人数字图书馆之间的互联,确立个人数字图书馆在信息检索中的角色与地位;(2)利用 P2P 关键技术对个人数字图书馆进行定位,发现个人数字图书馆资源列表;(3)利用 P2P 的信任机制,找到可信度高的个人数字图书馆,优化信息检索的质量。

3　Z39.50 与个人数字图书馆信息检索

P2P 主要针对 P2P 网络中不同节点上的文件名、

关键词等进行检索，这些信息一般是没有经过元数据的组织与管理的，也不需要实现特定的功能。通过元数据组织、加工整理后的个人数字图书馆中的信息资源（包括Web信息资源），要实现检索，同时实现个人数字图书馆与公共数字图书馆、个人数字图书馆与个人数字图书馆之间的互操作功能（下载元数据、并将自身的元数据上载到公共数字图书馆等），所有这一切都并非P2P技术所能实现的，客观上需要寻求具有上述功能的信息检索软件。目前，数字图书馆之间采用的主流信息检索协议Z39.50和基于Web服务的Z39.50的下一代新标准能充分实现以上功能。

2002年产生的ZING（Z39.50-International：Next Generation）——基于Web Service，XML，Soap等技术的下一代Z39.50检索协议集，其主要内容包括：简单的Web查询/获取协议（SRW/SRU）、通用检索语言（CQL）、面向对象的实现模型（ZOOM）、基于SOAP的现有Z39.50系统的转化机制（eZ3950）、如何检索与获取Z资源的详细信息等5个部分。SRW/U不是对Z39.50-1995版的更新和替代，而是在继承原有Z39.50标准合理成分的基础上建立的新的体系[10]。SRW/U的成熟和发展，最终不会简单地取代原有Z39.50标准，与原有Z39.50标准共同发展，在不同的领域发挥作用。随着Z39.50协议版本的不断升级，服务功能得到扩展，不再局限于原先图书馆界的书目文献信息，拓展到了整个Web信息资源。基于Z39.50协议的系统和模块已成为实现网上异构信息共享的理想工具。随着XML语言在网页设计、数据传输和文档编排方面的日益普及和应用，Web信息资源正逐渐以XML文档为主要表现格式，XML成为对文档结构进行编码和进行信息交换的基础标准，Z39.50协议强大的跨异构数据检索功能和XML元标记语言易于理解和使用的特性相结合[11]，使之成为主流检索协议。

4 基于P2P技术和Z39.50检索协议的个人数字图书馆检索策略

4.1 Z39.50在分布式个人数字图书馆信息检索中的不足

基于Z39.50协议的传统检索是Client/Server模式的，Client端送出一个查询请求，Server对Client端请求产生响应，并将检索结果处理后，返回给Client端。

理想的个人数字图书馆信息检索，既要使普通用户通过网络都能访问网络中分散的个人计算机中的个人数字图书馆信息资源，又能使个人数字图书馆用户能充分实现与个人数字图书馆之间及个人数字图书馆与公共数字图书馆之间的的互操作，以上功能均由Z39.50来实现。

然而，在个人数字图书馆信息检索过程中，个人数字图书馆呈分散状态，好似一盘散沙，无法对Client/Server定位；另外，网络中的个人数字图书馆缺乏唯一标识（服务器名称、IP地址及端口等），Z39.50的初始化机制、搜索机制、检索机制、访问控制机制等11种机制均无法在其检索中得到实施。需要采用一种互联、定位机制，形成动态的个人数字图书馆网，构建虚拟的Server，从而实现分布式检索功能。P2P技术能很好地解决这一难题。

4.2 实现P2P技术与Z39.50检索协议的有机结合

P2P技术与Z39.50检索协议的有机结合能给个人数字图书馆信息检索创造有力技术支撑。图2是一个基于P2P技术与Z39.50检索协议的个人数字图书馆检索策略模型。

4.2.1 目标端（Server）构建

Z39.50检索过程中的Server并非指协议本身的Server，而是指被检索信息资源本身，源端与目标端只需Z39.50协议，双方便可以实现通信。在P2P构建的个人数字图书馆网中，每一个对等点（个人数字图书馆）被检索时，它便可以成为Server。

（1）采用P2P技术进行定位、互联，将分散的个人数字图书馆构建成一个纯分散式个人数字图书馆P2P网络。PDLs-Peer为具有对等地位的节点（Peer），既可充当服务器，又可充当客户端双重角色，为成功实现个人数字图书馆信息资源检索提供了网络基础。

（2）纯分散式个人数字图书馆P2P网络中的一个个节点（Peer）在Z39.50协议（包括SRW/SRU）的作用下，形成具有纯分散式P2P结构的目标端（PDLs-Z-target），供用户检索。之所以要形成P2P网络，是因为能充分利用P2P信任机制[12]、P2P资源发现技术[13]，发现检索资源（包括个人数字图书馆节点IP、MARC地址、服务器名称、端口、数据库等）。

4.2.2 客户端配置（Client/Browser）

客户端，包括个人数字图书馆用户和非个人数字图书馆的普通用户，均须安装P2P软件，实现与目标端（PDLs-Peer）的连接。个人数字图书馆用户还得安装Z39.50检索协议；普通用户一般通过Browser访问

检索资源,因有 http-Z 网关的作用,一般不用安装 Z39.50 检索协议。建议个人数字图书馆软件系统必须具有 P2P 互联及带有 Z39.50 检索协议等功能。

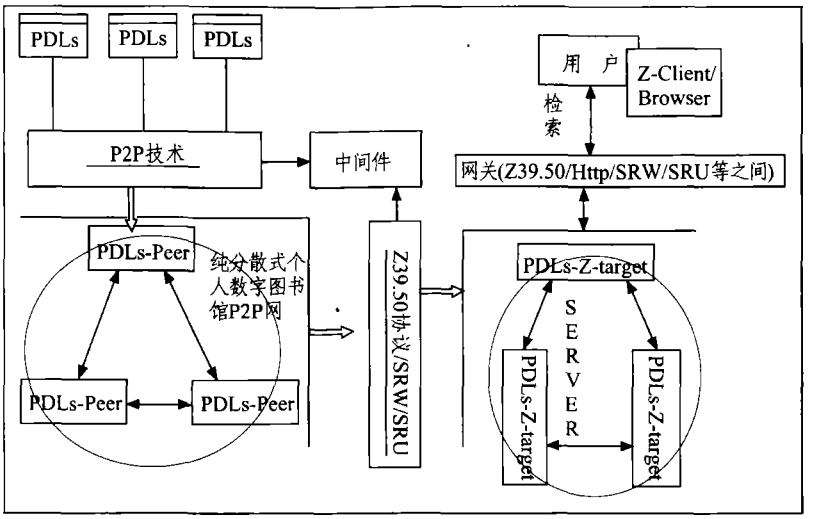

图 2　基于 P2P 技术与 Z39.50 检索协议的个人数字图书馆检索策略模型

4.2.3　*用户检索实现*

个人数字图书馆用户可通过 Z-Client、SRW/SRU-Z 网关,对 PDLs-Z-target 进行检索,并实现个人数字图书馆与公共数字图书馆、个人数字图书馆与个人数字图书馆之间的互操作。普通用户可以通过 Browser、http-Z 网关、http-SRW/SRU 网关,对个人数字图书馆中的元数据、Web 信息资源进行检索。

5　结论

P2P 技术和 Z39.50 协议在个人数字图书馆信息检索策略模型中有机结合,有助于充分实现分布式个人数字图书馆的检索、交互功能。构建 P2P 个人数字图书馆网络,实现个人数字图书馆信息检索,需要利用中间件技术将 P2P 技术、Z39.50 协议(包括 SRW/SRU)和个人数字图书馆软件捆绑在一起。下一步要研究的主要内容是中间件的开发、模型的验证和检索效率评价等。中间件直接关系到个人数字图书馆信息检索质量和互操作的顺利实现以及个人数字图书馆和个人知识管理的发展。

参考文献

1　陈光祚等. 论个人数字图书馆. 中国图书馆学报,2002(3)
2　臧国全. 适合于构建个人数字图书馆的全文数据库软件评析. 情报学报,2003(2)
3　张联峰等. 综述:对等网 P2P 技术. 计算机工程与应用,2003(12)
4　万淑超,金蓓弘,黄宇. P2P 平台的关键技术. 计算机科学,2005(6)
5　董健全,武雪丽,李智昕. P2P 网络中应用移动 Agent 进行资源搜索的研究. 计算机工程与设计,2005(1)
6　沈洁,胡金初. P2P 网络中的信息搜索技术. 福建电脑,2005(6)
7　姚佳丽,张坤龙,王珊. 基于 P2P 的数据索引与查询. 计算机科学,2005(3)
8　曹静霞,杨静,顾君忠. 基于推荐策略的 P2P 资源搜索算法研究与实现. 计算机应用,2005(8)
9　叶春,葛燧和,熊齐邦. 基于语义路由的 P2P 信息检索. 计算机仿真,2004(10)
10　于学锋,单启成. 下一代 Z39.50 技术探讨. 现代图书情报技术,2003(2)
11　胡一俊,焦玉英. Z39.50 和 XML 在信息获取中的应用. 现代图书情报技术,2003(5)
12　侯孟书等. P2P 系统的信任研究. 计算机科学,2005(4)
13　叶哲丽等. 基于 P2P 技术的资源发现机制的研究. 计算机工程与应用,2005(21)

张银犬　南通大学图书馆馆员,南京大学信息管理系博士生。通信地址:南京大学。邮编 210093。

朱庆华　南京大学信息管理系教授、博士生导师。通信地址同上。

(来稿时间:2006-09-21)