

- 徐树维^{1a,1b;2}, 齐惠颖^{1a,1b;3} (1. 中国科学院 a. 文献情报中心, b. 研究生院, 北京 100190; 2. 河南大学 计算中心, 河南 开封 475001; 3. 北京大学 医学部计算机教研室, 北京 100083)

P2P 在图书馆用户信息资源共享中的应用

[关键词] P2P; 信息资源共享; 用户信息资源

[摘要] 信息资源共享是 P2P 网络最典型的应用。本文提出的基于 P2P 的图书馆用户信息资源共享应用为用户间信息资源和知识的交流与共享提供了一种平台, 能最大程度地发挥信息资源的价值, 提高用户信息资源的利用率。

[中图分类号] G250.7

[文献标志码] A

[文章编号] 1005-8214(2009)10-0074-03

1 信息资源共享

信息资源共享是网络应用迅猛发展的主要驱动力, 而以计算机和网络为代表的现代信息技术为信息资源的共享提供了技术支持, 网络逐渐成为信息获取和信息交流的主要平台, 人们逐渐从对特定信息服务机构信息的依赖, 转向在用户信息需求的推动下如何快速、方便、容易、准确地获取与自己需求相关的信息资源, 他们不再关心信息为谁拥有、由谁提供这样的问题, 也不必了解信息的具体来源。^[1]

信息共享一方面扩大了可消费的信息资源的总和, 另一方面信息资源共享也是一种互利的行为, 拓展了信息用户的信息消费能力, 增加了信息福利。^[2] 信息资源共享的实质, 是信息资源应信息用户的需求在不同的信息服务提供方式之间的有序流动, 这种流动受到信息用户需求的牵引。^[1]

2 用户的信息资源

当前互联网上可用的海量数据信息几乎涵盖了人类活动的每一个方面, 并以非常快的速度在不断地增加。在这种情况下, 网络成为图书馆用户获取各种信息资源的主要途径。目前用户在使用网络资源时主要存在以下问题: 信息查找困难, 在互联网近乎无限的资源中很难有效地找到用户所需要的信息, 这就需要

用户在检索和处理信息的过程中对信息进行反复的检索、处理、组织和提炼; 信息获取困难, 只有用户才真正知道自己最需要的是什么, 这就需要用户花费大量的时间和精力对获取的信息按照自己关心的主题进行过滤、分析、整理, 最后形成满足用户需求的信息; 兴趣或研究方向相近的用户之间缺乏信息交流, 相关用户花费大量的时间和精力重复着上述的信息查找和信息获取的过程。

学术群体是数字图书馆最多、最重要的用户群体。^[3] 如果每类研究或兴趣相关的用户能把其他用户花费大量时间和精力查找和获取的信息资源视为自己信息资源的一种扩展, 反之自己的资源也是其它用户或群体资源的扩展, 就可以提高图书馆用户在科研和学习过程中的信息获取效率, 实现资源利用价值的最大化。

3 用户间信息资源共享

用户间信息资源共享方式传统上是通过用户间直接对话、书信往来、交换出版物等这些非正式学术交流形式来获取所需要的学术信息资源, 共享的范围极为有限, 很多情况下很难满足用户的需求。现在通过网络, 用户间信息传递迅速, 针对性强, 选择性广, 但信息缺乏有效的质量监督, 可靠性较低。

传统上图书馆曾有人们所需要的最重要的信息资源, 现在也是科研用户获取信息的主要途径之一, 因此图书馆能在用户间信息资源共享中起着重要的作用, 为用户之间信息资源的共享提供一种中介。这样信息资源就不再是在图书馆与用户之间单向流动, 而是在用户和用户之间、用户和图书馆之间双向流动, 图书馆就可以在网络信息环境中, 主动为用户提供服务, 成为图书馆用户之间进行信息资源交流和共享的平台。

4 P2P 技术

P2P 的全称为“Peer-to-Peer”, 即对等互联网络技术(点对点网络技术), 其目的是使任何网络设备可以为其他网络设备提供服务, 系统中的任何一个

节点 (peer) 之间都能通过直接交换信息来进行信息和服务的共享,而不需要经过其它的中间实体。

4.1 P2P 的概念、特点及模型

P2P 是一种分散的、分布式的资源管理模型,系统中的任何一个节点 (peer) 之间都能通过直接交换信息来进行信息和服务的共享,它最根本的思想是网络中的节点既可以获取其它节点的资源或服务,同时又是资源或服务的提供者,即兼具客户机和服务器双重身份,用户可以直接控制和共享资源,这使得用户能以更方便快捷的方式参与到信息资源的共享中。

P2P 技术最典型的应用是信息资源的共享与存储,其有如下特点^[4]: (1) 对等。每个节点具有相同的地位,既是资源提供者又是资源消费者,同时扮演着 C/S 模式中的服务器和客户机两个角色。(2) 自治。节点可以独立自主地决定共享什么资源,允许谁共享以及在什么条件下进行共享,节点可以随时自主地加入或退出系统。(3) 动态。节点自身也可以随时加入或离开网络,因此整个 P2P 网络是动态的。(4) 无集中控制和大规模分布。传统的分布式计算系统中一般存在集中式的控制,而在 P2P 系统中,这样的控制不存在。(5) 系统自组织。自组织定义为系统组织的自发增加过程,即这种增加不受环境或其它外部系统的控制。

由于各节点之间可以直接交互共享资源,资源利用率得到极大提高,因而在 P2P 平台上进行信息资源的共享是一种较理想的方案,可以有效利用 P2P 本身的优势高效,最大程度地实现信息共享。

P2P 模型从拓扑结构上可分为: (1) 集中模型 (Napster): 采用一个集中服务器专门提供服务,维护简单,发现算法灵活高效并能够实现复杂查询,是目前使用最普遍的网络拓扑结构,如下图 (a); (2) 分散模型 (Gnutella): 采用非规则的结构,整个网络的拓扑结构是随机的和不规则的,面对网络的动态变化具有较好的容错能力,但由于没有确定的拓扑结构,无法保证资源发现的效率,如下图 (b); (3) 层次化模型 (Mbone): 吸取了集中模型和分散模型的优点,采用树型拓扑结构,性能可扩展性较好,较容易管理,但对超级节点依赖程度高,如下图 (c); (4) 混合模型 (Edonkey): 把整个 P2P 网络建成一个二层

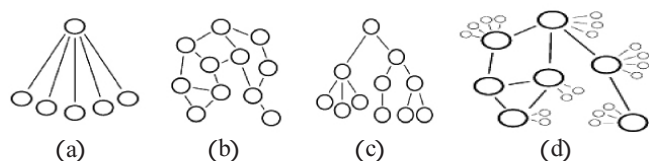


图 P2P 模型

结构,由普通节点和超级节点组成,一个超级节点管理多个普通节点,超级节点构成完全分散式模型,如下图 (d)。

4.2 基于 P2P 数字图书馆的研究

目前基于 P2P 的数字图书馆系统研究主要集中在基于 P2P 架构的数字图书馆原型研究、异构模式互操作研究和元数据整合研究。^[5,6] 在基于 P2P 的数字图书馆原型研究方面,英国兰开斯特大学的 P2P-4-DL 项目建立了一个基于 P2P 架构的数字图书馆系统,该系统采用了半集中式网络结构,通过提供一组一般的面向应用的服务,使得用户可以更加有效地利用网络资源,并免去用户理解底层 P2P 技术的负担。^[7] 文献 [8] 提出一个高效的结构化 P2P 的数字图书馆馆藏搜索体系结构和新的基于关键词的索引策略,显著地改进了查询性能。还有基于 P2P 的信息检索、^[9-11] 馆际合作研究,^[12] 以及基于 P2P 的多媒体数字图书馆,^[13] 基于 P2P 的个人数字图书馆社区^[14] 的应用等。

5 基于 P2P 的图书馆用户间信息资源的共享

每个用户个人计算机上都有许多有价值的各种信息资源能与其他用户共享。P2P 技术和数字图书馆技术的新发展给我们启示去寻找一种新的方式来实现用户间信息资源的共享。基于 P2P 的图书馆用户资源共享致力于通过 P2P 架构来构建一个环境,使用户能在这个平台上发布、检索和下载自己所需要的各种信息资源。在基于 P2P 的信息共享环境中,每个节点都是主动参与者,都提供一定的资源,随着加入节点的增多,系统能提供的资源越来越丰富。

5.1 P2P 系统架构

P2P 的自治性的要求使得信息共享环境中不存在中央管理机制,节点之间拥有对等的关系,这为系统带来了自组织,容错性好,可扩展性强等优点,但由于节点的自治以及节点信息的动态变化,每个节点只知道自己局部的拓扑信息,不能一次性构建出高产的覆盖网络拓扑,也不能构建永远高效的覆盖网络拓扑。另外考虑到科研或专业用户从图书馆获取的电子资源都有相应的知识产权信息,因此在基于 P2P 的图书馆用户间信息资源共享应用的设计中,基础架构采用的是一个集中模型,采用集中式检索方法,所有 P2P 节点都与已知地址的 P2P 目录服务器相连,服务器负责对 P2P 网络中的共享文件进行索引和查询,服务器集中存放对等节点的地址信息和数据信息。另外,有服务器存在用户自然会感到安全,服务器可采用授权访问来保护用户和共享平台免受恶意入侵。

5.2 P2P 目录服务器

P2P 信息资源共享平台通过一个中央服务器保存所有用户共享的文件索引和文件存放位置的信息,这样便于系统进行用户和网络流量的管理。用户所共享的资源列表都存放在服务器的数据库中,服务器保留这些资源的详细描述,在用户之间初始化一个链接,为访问的用户列出所需内容的文件夹或者目录。当用户从目录服务器上检索到自己需要的信息时,可以根据目录服务器上共享信息资源节点的地址和数据信息,找到存放共享信息资源的个人计算机,并建立链接,直接进行共享资源的传输。目录服务器用于辅助用户节点之间建立链接,一旦链接成功,对等节点之间即可直接进行通信。这样就实现了文件查询和文件传输的分离,有效地节省了中央服务器的带宽,减少了系统的文件传输延时。另外服务器维护了各个节点的名字、IP 地址和节点提供的共享资源的列表,当节点加入或离开该系统时,P2P 目录服务器会随之更新系统索引表。

5.3 P2P 用户节点

用户节点的客户端程序有两个,一个用于登录,一个用于浏览。登录程序的主要作用是登录到服务器上,告诉服务器自己所共享的资源列表和节点信息,同时来自别的节点请求也通过它和用户之间建立链接,处理请求。浏览程序的主要作用是浏览、检索目录服务器上共享的信息资源信息,然后根据目录服务器上的地址和资源信息,在节点之间直接建立链接,实现信息资源的共享。

5.4 P2P 用户社区

P2P 用户社区的基本思想是将研究或兴趣相近的用户划分到同一个社区中,当用户需要某类信息时,首先考虑在社区内查找,可根据用户兴趣提供个性化信息检索服务,以提高信息检索效率。

图书馆用户因专业、爱好和研究兴趣的不同,所关注的信息资源类型也有所差异,因此在以图书馆为中介的 P2P 信息资源共享模型中,完全可以根据用户的类别信息将 P2P 共享空间按专业、爱好和兴趣划分为不同的社区,同一社区由具有兴趣、爱好相同或相近的节点组成,用户也可以主动选择加入一个或多个社区。

当用户下载共享信息资源的时候,从某种程度上就可以认为上传共享信息资源的用户和下载相应共享信息资源的用户之间存在某种关联,可以对目录服务器上的日志信息进行统计分析,当用户间这种关联达

到某种阈值的时候,就可以自动根据这种相关性为这些用户建立相应的兴趣组或用户社区,建立一种相互影响的学习或研究关系,最大程度地发挥用户所获取的信息的价值,使研究或兴趣相关及相近的用户之间在信息获取与交流方面得到更多的便利。

兴趣相近的节点存放的内容和提交的检索查询也相近。可以通过挖掘节点所共享的内容所反应的节点兴趣,从而使用户在需要时能够高效地检索到自己所需要的内容。通过用户聚类,可以根据相似用户来推荐资源,或根据当前下载情况,预测当前用户的兴趣,为当前用户提供 k 个最相似的邻居。^[15]

通过对信息资源的重组,使兴趣相同的用户节点相互合作,简化信息资源的存储、查找和使用,从而提高资源利用率。此外,人类社会的聚群心理和聚群行为也反映到 P2P 系统中,体现为节点具有利人和利己的双重动机,一方面提供共享资源,另一方面也有自身的资源需求。^[4]

6 结语

本文中提出的基于 P2P 的图书馆用户间信息资源的共享应用为爱好、兴趣和研究相关或相近的用户提供了一种信息资源和知识的交流平台,使用户能以较低的成本及较快的速度,方便、容易地实现信息资源和知识的共享与交流。虽然 P2P 在信息共享中的应用还要解决知识产权、网络安全等方面的问题,但在图书馆网络环境内实现各种资源的共享还是可行的。

[参考文献]

- [1] 凌美秀. 从有限到无限: 信息资源共享的演变路径 [J]. 图书情报知识, 2007 (116): 72-75.
- [2] 马费成, 裴雷. 信息资源共享及其效率分析 [J]. 情报科学, 2004, 22 (1): 1-8.
- [3] N Meyyappan, etc. Design and Development of a User-Centred Digital Library System: Some Basic Guidelines [EB/OL]. [2008-04-15]. <http://ncsi-net.ncsi.iisc.ernet.in/gsd/collect/icco/index/assoc/HASH011e.dir/doc.doc>.
- [4] 唐九阳. 基于对等计算的信息共享相关技术研究 [D]. 长沙: 国防科学技术大学, 2006.
- [5] 李广建, 龚立群. 下一代数字图书馆系统体系架构的发展动向 [J]. 高校图书馆工作, 2006, 26 (133): 1-5.
- [6] Benjamin, etc. OAI-P2P: A Peer-to-Peer Network for Open Archives [EB/OL]. [2008- (下转第 93 页)]

询过程中存在的问题,笔者下一步将继续探索实时咨询效率测度的方法,致力于改善实时咨询服务效果,提升服务品质。

[参考文献]

- [1] Coffman S. Arret L- To chat or not to chat- taking another look at virtual reference [J]. *Searcher*, 2004 (12).
- [2] 袁红军. 虚拟参考咨询服务效率分析 [J]. *中国图书馆学报*, 2006 (4): 47-49.
- [3] 许激. 效率管理——现代管理理论的统一 [M]. 北京: 经济管理出版社, 2004.
- [4] Marie L, etc. A Multiple- case study investigation of the discontinuation of nine chat reference services [J]. *Library & Information Science Research*, 2006, 28 (4): 521-547.
- [5] Lankes RD, etc. Cost, statistics, measures, and stan-

dards for digital reference services :A preliminary view [J]. *LibraryTrends*, 2003 (5): 401-413.

- [6] 中国科学院国家科学图书馆 [EB/OL]. [2009-03-29]. <http://www.las.ac.cn/las/service/intro.jsp>.
- [7] 梁南燕, 等. 数字参考咨询服务的质量与效益评估 [J]. *图书情报工作*, 2005, 49 (8): 121-123.
- [8] (美) 麦克卢尔, 等. 数字参考咨询服务的质量评价 [M]. 卢海燕, 王磊译. 北京: 北京图书馆出版社, 2007: 5-6.

[作者简介] 李娟娟 (1984-), 女, 中国科学院国家科学图书馆硕士研究生, 发表论文 1 篇; 周宁丽 (1953-), 女, 中国科学院国家科学图书馆研究馆员, 硕士生导师, 发表论文 20 余篇。

[收稿日期] 2009-06-05 [责任编辑] 肖群

(上接第 76 页) 04-17]. <http://ieeexplore.ieee.org/iel5/8062/22290/01039766.pdf?arnumber=1039766>.

- [7] J Walkerdin, P Rayson. P2P- 4- DL: Digital Library over Peer- to- Peer [EB/OL]. [2008-04-17]. csdl.computer.org/comp/proceedings/p2p/2004/2156/00/21560264.pdf.
- [8] Ivana Podnar, etc. A Peer- to- Peer Architecture for Information Retrieval Across Digital Library Collections [EB/OL]. [2008-04-17]. <http://nlp.uned.es/~julio/ECDL- best- paper/552.pdf>.
- [9] Matthias Bender, etc. Towards Collaborative Search in Digital Libraries Using Peer- to- Peer Technology [EB/OL]. [2008-04-17]. <http://www.minerva-project.org/publications/DELOS04.pdf>.
- [10] Yan Chen, etc. A Scalable Semantic Indexing Framework for Peer- to- Peer Information Retrieval [EB/OL]. [2008-04-17]. <http://hdir2005.isti.cnr.it/camera-ready/7.Chen.pdf>.
- [11] Wolf- Tilo Balke, etc. DL meets P2P- Distributed Document Retrieval based on Classification and Content [EB/OL]. [2008-04-17]. <http://www.kbs.uni-hannover.de/Arbeiten/Publikationen/2005/p2pdl- eccl 2005.pdf>.
- [12] WEN- CHUNG SHIH, etc. An Interlibrary Cooperation Framework for Digital Libraries Using P2P Technology [EB/OL]. [2008-04-17]. [\[zona.edu/1375/01/22.Wen- Chung_Sih_pp156- 159_.pdf\]\(http://www.kbs.uni-hannover.de/Arbeiten/Publikationen/2005/p2pdl- eccl 2005.pdf\).](http://dlist.sir.ari-</p>
</div>
<div data-bbox=)

- [13] Yi Ma, Ramazan S Aygiin. Peer- to- Peer Based Multimedia Digital Library [EB/OL]. [2008-04-15]. ieeexplore.ieee.org/iel5/9879/31411/01460891.pdf?arnumber=1460891.
- [14] Yanfei X U. A P2P Based Personal Digital Library for Community [EB/OL]. [2008-04-17]. <http://ieeexplore.ieee.org/iel5/10544/33357/01579033.pdf>.
- [15] 曾春, 等. 个性化服务技术综述 [J]. *软件学报*, 2002, 13 (10): 1952-1961.

[作者简介] 徐树维 (1969-), 男, 中国科学院国家科学图书馆情报学专业网络信息管理技术与信息系统研究方向在读博士生; 齐惠颖, 女, 中国科学院国家科学图书馆情报学专业网络信息管理技术与信息系统研究方向在读博士生。

[收稿日期] 2009-02-20 [责任编辑] 肖群