

## P2P 情境下基于用户兴趣的个人数字图书馆资源定位

张红伟,任淑敏

**[摘要]**根据个人数字图书馆的分散性特征,将 P2P(Peer to Peer)技术应用于个人数字图书馆系统,提出一种基于 P2P 技术的个人数字图书馆共享模型-MIPDLM,以提高个人数字图书馆共享水平。在该模型中,个人数字图书馆可根据兴趣分簇,探讨拓扑重构的过程,其基于多兴趣的搜索机制,可将搜索消息不断向兴趣相似度较高的个人数字图书馆转发,因此能大大缩小搜索范围。仿真结果表明,与随机漫步(Random Walk)机制相比,MIPDLM 模型可在一定程度上提高个人数字图书馆搜索成功率,减少平均路径长度,提高资源定位效率。

**[关键词]**个人数字图书馆;P2P;兴趣;搜索

**[中图分类号]**G250.76;R-058

**[文献标志码]**A

**[文章编号]**1671-3982(2017)11-0030-06

### Location of resources in personal digital library with P2P technology based on interests of users

ZHANG Hong-wei, REN Shu-min

(Library of Jining Medical College, Jining 272067, Shandong Province, China)

Corresponding author:REN Shu-min

**[Abstract]**A P2P technology-based personal digital library sharing model, MIPDLM, was established using the P2P technology according to the decentralized characteristics of personal digital library in order to improve the personal digital library sharing, in which the topology remodeling was studied according to the interest clusters in personal digital library. The searched messages are constantly transmitted to personal digital library with a high interest similarity, which can thus reduce the searching scope. Simulation experiment showed that MIPDLM could improve the successful searching rate, reduce the average path length, and increase the location rate of resources.

**[Key words]**Personal digital library; P2P; Interest; Searching

近年来,个人计算机上存储的个性化信息资源数量大幅增长,有效管理和使用这些资源成为亟待解决的问题,因此个人数字图书馆应运而生。目前个人数字图书馆有两种形式:一种是基于服务器的

个人信息空间,一般称作“My library”(我的图书馆)<sup>[1]</sup>;另一种是基于 PC 的数字信息资源库,一般称作“Personal Digital Libraries”(PDLs)<sup>[2]</sup>。本文研究的是第二种形式。

作为个人信息工具,个人数字图书馆的职能主要有两种:一种是有效管理和使用个人计算机上的信息资源,这是个人数字图书馆最基本的职能;另一种是共享个人计算机上的信息资源,促进资源的有效利用和知识再生,这是个人数字图书馆的核心功能。信息搜索是资源共享的基础,没有合适的信息搜索机制就无法构建高效的资源共享平台。由于个人数字图书馆的分散性特点,其信息搜索及共享研究在传统的客户机/服务器模式下难以突破。为了

**[基金项目]**山东省高等学校人文社会科学计划项目“基于对等网络的个人数字图书馆关键技术研究”(J16WJ08);山东省档案局科技项目“基于 P2P 网络技术的档案共享平台模型研究”(2015-05)

**[作者单位]**济宁医学院图书馆,山东 济宁 272067

**[作者简介]**张红伟(1981-),男,山东济宁人,硕士,副研究馆员,发表论文 10 篇。

**[通讯作者]**任淑敏(1963-),女,山东济宁人,教授,馆长,发表论文 40 余篇。

解决这个问题, P2P (Peer to Peer) 技术被引入了个人数字图书馆领域。

## 1 个人数字图书馆与 P2P 技术简介

大量学者对个人数字图书馆的资源共享进行了研究。王小立指出资源共享是个人数字图书馆的本质属性, 其发展仍处于初始阶段, 具有广阔的前景<sup>[3]</sup>; 陈春艳分析了阻碍个人数字图书馆资源共享的因素, 指出传统的集中式模式(客户机/服务器和浏览器/服务器)阻碍了个人数字图书馆资源共享的发展<sup>[4]</sup>; 张哲对比了集中式和 P2P 两种模式, 指出集中式模式共享成本高、效率低, 不适合个人数字图书馆资源共享的发展, 而 P2P 模式的去中心化与个人数字图书馆的分散性不谋而合, 不需要服务器的参与, 能充分发挥个人数字图书馆的能力, 适合个人数字图书馆的发展<sup>[5]</sup>。

P2P 即对等网络, 是相对传统的客户机/服务器模式而言的。系统中每个节点的作用都是对等的, 既可以充当服务器为其他节点提供资源和服务, 也可以充当客户机享用其他节点的资源和服务。P2P 技术因其良好的自组织能力和可扩展性, 而受到业界的广泛关注, 主要应用于搜索引擎、资源共享、协同工作和实时通讯等领域。目前, P2P 技术在个人数字图书馆方面的应用研究较少, 主要集中于文件共享模型的构建和信息搜索。王春梅等介绍了个人数字图书馆的共享现状, 尝试用 P2P 技术构建简单的个人数字图书馆共享架构, 并通过模拟实验验证了 P2P 应用于个人数字图书馆的可行性<sup>[6]</sup>; 厦门大学 Xu Yanfei 利用具有相同兴趣的个人数字图书馆形成社区, 探讨了适合个人数字图书馆的 P2P 网络拓扑结构<sup>[7]</sup>。资源定位是 P2P 技术应用于个人数字图书馆需要突破的一个难题。张银犬等指出在分布式环境下个人数字图书馆的信息检索还是一个研究空白和难点, 并结合 P2P 技术和 Z39.50 协议, 对个人数字图书馆信息检索策略进行了初步探索<sup>[8]</sup>。

分布式环境下没有集中式环境中掌管全局信息的中央服务器, 资源定位成为一大难题。在非结构化 P2P 网络中, 洪泛是最常使用的一种检索策略。这种策略将查询消息向所有邻居进行广播, 造成大量通讯开销, 严重影响了搜索效率<sup>[9]</sup>。为了减少消息数量, 一些学者提出了启发式搜索机制, 利用搜索

历史、节点状态等信息, 将消息转发范围控制在局部存有目标资源几率较高的节点, 可缩小搜索范围, 提高搜索效率。基于兴趣的搜索机制就是启发式搜索中的一种<sup>[10]</sup>。文献[11]根据节点之间的交流历史创建兴趣捷径, 从而改进 Gnutella 的搜索性能。何可等采用基于兴趣的完全二叉树 P2P 拓扑结构, 提出了一种基于 Flooding 机制的双向资源搜索算法<sup>[12]</sup>; 陈香香等根据节点共享文件的差异将网络分域, 并且利用预算值控制搜索范围<sup>[13]</sup>。以上研究将最主要的兴趣作为节点的唯一兴趣, 但是节点往往有多个兴趣, 当节点搜索其他兴趣的内容时, 就无法体现基于兴趣搜索机制的优势。

在以上研究的基础上, 针对个人数字图书馆搜索机制尚不完善的问题, 本文提出一种基于 P2P 技术采用多兴趣的搜索机制的个人数字图书馆模型 (Multi-Interest Personal Digital Library Model, MIPDLM)。

## 2 基于 P2P 技术的个人数字图书馆模型 MIPDLM

### 2.1 兴趣的表示

作为个性化特征突出的个人信息工具, 个人数字图书馆(PDL)所存储的资源(本文只研究文档类的资源)往往表现出一定的兴趣偏好。传统的兴趣表示通常只关注 PDL 最主要的兴趣, 这种单一的、静态的兴趣表示方法无法兼顾 PDLs 多样的、动态变化的兴趣, 导致 PDL 搜索效率低下。本文采用多兴趣表示方法, PDL 可以加入多个兴趣簇, 即使 PDL 拥有丰富的动态变化的兴趣, 系统也能将搜索消息转发到相似度较高的兴趣簇, 故可大大缩小搜索范围, 提高搜索效率。

本文采用经典的向量空间模型 (Vector Space Model, VSM) 作为兴趣表示模型, VSM 模型将现实中无法相互比较的兴趣转化为可以计算彼此相似度的向量, 如  $PDL_i$  的兴趣向量为  $\{PDL_{i1}, PDL_{i2}, \dots, PDL_{ik}, \dots, PDL_{in}\}$ , 表示  $PDL_i$  共有  $n$  个兴趣, 其中  $PDL_{ik}$  是  $PDL_i$  中第  $k$  ( $1 \leq k \leq n$ ) 个兴趣向量。两个兴趣之间的相似度采用夹角余弦表示, 夹角余弦值越大, 表示两个兴趣之间的相似度越大; 反之, 两个兴趣之间的相似度越小。同理, 搜索消息向量与文档向量之间的兴趣相似度也取其夹角余弦值。兴趣与 PDL 之间的相似度是指该兴趣与 PDL 中所有兴

趣相似度的最大值。同理,搜索消息向量与 PDL 之间的兴趣相似度也是如此计算。

### 2.2 兴趣覆盖网络的构建

构建适合个人数字图书馆资源共享的兴趣覆盖网络。在 MIPDLM 模型动态构建过程中,系统需用到个人数字图书馆的一些数据信息。每个  $PDL_i$  存储如下 3 方面的信息:一是  $PDL_i$  共享文档的集合,二是  $PDL_i$  的兴趣向量集合,三是两种邻居集合即簇内邻居集合和簇外邻居集合。簇内邻居集合存储与  $PDL_i$  兴趣相似度较高的 PDLs,簇外邻居集合存储与  $PDL_i$  兴趣相似度较低的 PDLs。簇内邻居和簇外邻居都包含一组  $(ip(PDL_j), PDL_{jr})$  信息,其中簇内邻居集合中  $ip(PDL_j)$  是  $PDL_i$  簇内邻居  $PDL_j$  的 IP 地址, $PDL_{jr}$  是与  $PDL_i$  的第  $k$  个兴趣相对应的兴趣向量。相反,簇外邻居集合中  $ip(PDL_j)$  是  $PDL_i$  的簇外邻居  $PDL_j$  的 IP 地址, $PDL_{jr}$  是与  $PDL_i$  的第  $k$  个兴趣相对应的兴趣向量。大量的 PDLs 根据簇内邻居形成一个个兴趣簇,这些兴趣簇紧密地结合在一起组成一个兴趣覆盖网,这种网络类似于一个规则网络。为了保证“小世界”特性,PDLs 通过少量的簇外邻居链接到其他的兴趣簇<sup>[14]</sup>。

本文采用全分布式非结构化的 P2P 网络结构。当  $PDL_i$  加入网络时,首先初始化兴趣向量  $\{PDL_{i1}, PDL_{i2}, \dots, PDL_{ik}, \dots, PDL_{im}\}$ ,针对每个兴趣  $PDL_{ik}$ , $PDL_i$  维护两个路由表  $CRPDL_{ik}$  和  $NCRPDL_{ik}$ , $CRPDL_{ik}$  存储簇内邻居信息, $NCRPDL_{ik}$  存储簇外邻居信息。 $PDL_i$  利用随机漫步算法在网络中发出一个消息,选取收到消息的  $2m$  个 PDLs 作为初始值,分别将  $m$  个 PDLs 的 IP 和兴趣向量存入  $CRPDL_{ik}$  和  $NCRPDL_{ik}$ 。PDLs 的兴趣一直处于动态变化过程中,为了确保网络拓扑结构适应个人数字图书馆兴趣的变化,使兴趣相似度较高的 PDLs 始终聚集在一起,PDL 定期检测每个兴趣  $PDL_{ik}$  的聚合度  $DPDL_{ik}$ 。聚合度  $DPDL_{ik}$  的大小是  $PDL_i$  与所有  $PDL_{ik}$  相应的簇内邻居兴趣相似度的平均值,如果  $DPDL_{ik}$  高于或等于某一临界值  $\gamma$ ,说明当前  $PDL_i$  与簇内邻居兴趣相似度较高, $PDL_i$  不需要搜寻新的簇内邻居;如果  $DPDL_{ik}$  低于某一临界值  $\gamma$ ,说明当前  $PDL_i$  与簇内邻居关于  $PDL_{ik}$  的兴趣相似度较低,系统会重新搜寻兴趣相似度较高的 PDLs 并加入  $CRPDL_{ik}$  中。

### 2.3 基于多兴趣的搜索机制

哈佛大学社会心理学教授 Stanley Milgram 在著名的连锁信实验中发现了典型的小世界现象。受该实验的启发,在搜索过程中,PDL 将消息转发到与搜索信息兴趣相似度较高的 PDL 中,这些通过兴趣连接在一起的 PDLs 不停地将消息转发到它们认为最有可能拥有目标资源的 PDLs 中,可大幅度减少搜索范围。

基于多兴趣的算法流程见图 1。其搜索过程如下:步骤①PDL<sub>i</sub> 发起一个搜索消息  $q$  时,首先将  $q$  的搜索内容初始化为向量  $d$ ,定义  $q$  的最大转发跳数  $TTL_q$ ,初始化  $q$  所经过的 PDLs 列表 ADD 和搜索到的文档列表 DOC。其中,ADD 存放消息  $q$  所经过的 PDLs 的 IP,DOC 存放搜索到的文档向量以及相应 PDL 的 IP。计算 PDL<sub>i</sub> 中与  $d$  相似度最大的兴趣 PDL<sub>ik</sub>,在路由表  $CRPDL_{ik}$  中计算与  $d$  兴趣相似度最大的兴趣 PDL<sub>jr</sub>。步骤②将搜索消息  $q$  转发到 PDL<sub>j</sub>,把 PDL<sub>j</sub> 的 IP 添加到 ADD 中,计算  $d$  与 PDL<sub>j</sub> 中所有的共享文档的兴趣相似度。当高于阈值  $\theta$  时,把该文档向量和 IP 添加到文档列表 DOC 中,将  $tTL_q$  减 1。如果  $TTL_q$  等于 0,搜索消息  $q$  返回 PDL<sub>i</sub>,否则转到步骤③。步骤③计算  $CRPDL_{jr}$  中与  $d$  相似度最大的兴趣 PDL<sub>lm</sub>。如果列表 ADD 中含有 PDL<sub>i</sub> 的 IP,说明消息  $q$  已经到达过 PDL<sub>i</sub>。为了防止环路的产生,舍弃 PDL<sub>i</sub>,按兴趣相似度由大到小的顺序继续在  $CRPDL_{jr}$  中寻找,直到找到 IP 不在 ADD 中的 PDL<sub>lm</sub>(如果  $CRPDL_{jr}$  和  $NCRPDL_{jr}$  中所有 PDLs 的 IP 都在列表 ADD 中,则搜索结束,消息  $q$  返回 PDL<sub>i</sub>),其兴趣相似度值为  $value$ 。如果  $value$  小于阈值  $\lambda$ ,表明  $CRPDL_{jr}$  中的 PDLs 与  $d$  兴趣相似度不高。在  $NCRPDL_{jr}$  中搜寻与  $d$  相似度最大的兴趣 PDL<sub>sn</sub>(检查相应的 IP 是否在列表 ADD 中,如果在列表 ADD 中则利用上面的方法寻找下一个兴趣),从 PDL<sub>lm</sub> 与 PDL<sub>sn</sub> 中选取与  $d$  兴趣相似度较大的那个兴趣作为 PDL<sub>jr</sub>。如果  $value$  大于或等于阈值  $\lambda$ ,表明  $CRPDL_{jr}$  中的 PDLs 与  $d$  兴趣相似度较高,不必再去  $NCRPDL_{jr}$  中寻找相似度更高的兴趣而将 PDL<sub>lm</sub> 作为新的 PDL<sub>jr</sub>,重复步骤②。当消息返回到 PDL<sub>i</sub> 时,搜索到的文档将按与  $d$  兴趣相似度由大到小的顺序显示出来。

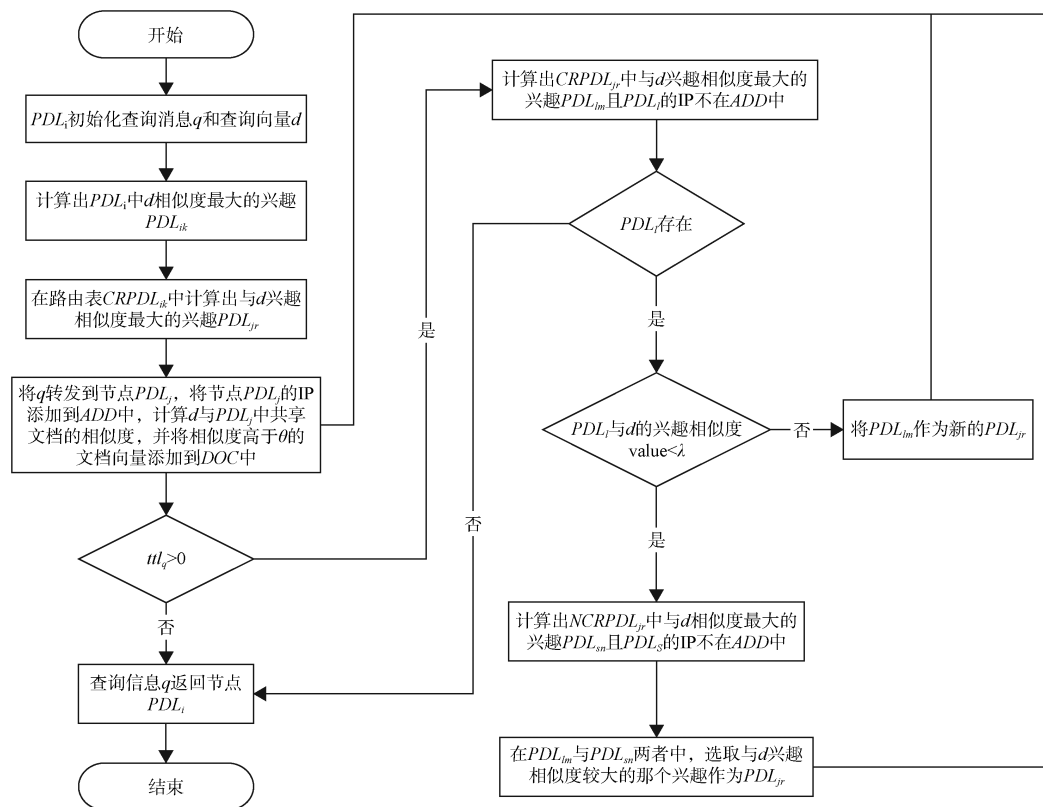


图 1 个人数字图书馆基于多兴趣的资源定位算法流程

### 3 仿真实验

#### 3.1 实验设置

从中国知网期刊论文数据库中选取实验文档。根据中国知网期刊论文数据库的学科分类,从基础科学、经济与管理科学等 10 个学科分类中分别选取 20 个兴趣主题,每个兴趣主题下载 50 篇论文。实验工具采用 P2P 通用网络模拟器 PeerSim。

在 PeerSim 中生成一个网络规模为 1 000 个 PDLs 的全分布式非结构化 P2P 网络,将选取的 10 000 份文档散布在 1 000 个 PDLs 中,每个 PDL 的文档所涉及的兴趣数量不超过 3 个。

#### 3.2 兴趣聚合度的确定

系统开始运行后,每个 PDL 都尝试启动聚合度检测程序,寻找兴趣相似度较高的 PDLs。为了使 PDLs 不在同一时刻启动检测程序,所有 PDLs 在 0-30 分钟内随机选择某一时刻启动检测程序。第一次启动检测程序后,系统设定 PDL 定期启动一次检测程序,运行 10 个周期,计算搜索成功率。兴趣相似度较高的 PDLs 聚集在一起是 MIPDLM 模型正常

运行的基础。阈值  $\gamma$  决定最小聚合度,选取合适的  $\gamma$  值是 MIPDLM 模型提高搜索效率的前提。为了验证阈值  $\gamma$  如何取值合适,实验测试了不同  $\gamma$  值下的搜索成功率,实验重复 10 次,每次系统都重新初始化网络,为每个 PDL 分配文档, TTL = 6,实验结果取平均值(图 2)

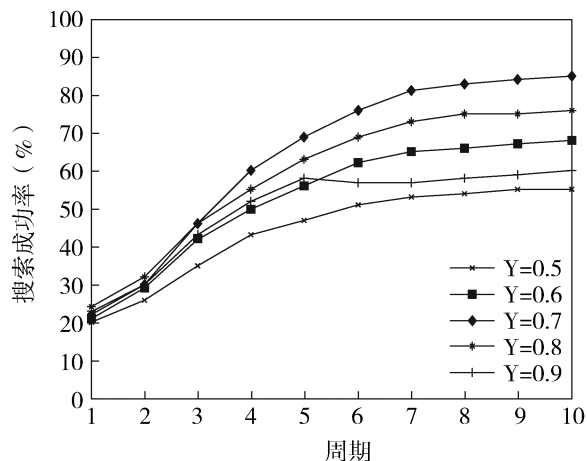


图 2 不同  $\gamma$  值下的搜索成功率

从图 2 可以看出,当  $\gamma=0.5$  和  $\gamma=0.6$  时,聚合度太低,兴趣相似度较高的 PDLs 没能聚集在一起,搜索成功率较低;当  $\gamma=0.8$  和  $\gamma=0.9$  时,兴趣相似度较高的 PDLs 虽然聚集在一起,但是聚合度太高,节点无法通过兴趣相似度较低的 PDLs 搜索资源,搜索成功率仍然不高;当  $\gamma=0.7$  时,聚合度较为合适,兴趣相似度较高的 PDLs 聚集在一起,搜索效果较为理想。以下实验中  $\gamma$  值均取 0.7。

### 3.3 实验结果及分析

搜索机制的优劣主要取决于搜索成功率和平均路径长度等因素。搜索成功率是指搜索成功次数在搜索总次数中所占的比重,搜索成功率越高搜索性能越好;平均路径长度是指搜索路径长度的平均值,平均路径长度越长需要搜索的 PDLs 越多,在增加延迟的同时也加重了网络负载,较低的平均路径长度是搜索机制追求的重要目标。

随机漫步(Random Walk)是 P2P 网络中性能较为优越的一种搜索机制<sup>[15]</sup>。为了验证 MIPDLM 模型的性能,需对 MIPDLM 模型中搜索成功率和平均路径长度与 Random Walk 机制进行比较。MIPDLM 网络模型运行一段时间,兴趣簇达到一个相对稳定的阶段后,随机选取 PDL 根据自己的兴趣进行搜索, *TTL* 从 1 增加到 10,每个 *TTL* 搜索 200 次,计算结果平均值。

搜索成功率如图 3 所示。在 *TTL* 相同的情况下, MIPDLM 模型的搜索成功率明显高于 Random Walk 的。随着 *TTL* 不断增加, Random Walk 的搜索成功率一直处于缓慢上升的趋势,并且在 *TTL*<4 之前搜索成功率上升的幅度较小。当 *TTL* = 10 时, Random Walk 的搜索成功率只有 80% 左右;当 *TTL* = 4 时, MIPDLM 模型的搜索成功率已接近 80%,说明 MIPDLM 模型中由于 PDLs 形成兴趣簇,兴趣一致的 PDLs 聚集在一起,搜索消息转发到兴趣相似度较高的 PDL 中,可大大缩小搜索范围,绝大部分 PDLs 能在 4 跳内搜索到所需资源。搜索平均路径长度如图 4 所示。MIPDLM 模型的搜索平均路径长度明显低于 Random Walk 的。Random Walk 的平均路径长度随着 *TTL* 的增加不断攀升。当 *TTL* = 10 时, Random Walk 的平均路径长度已接近 7。随着 *TTL* 的增加, MIPDLM 模型的平均路径长度维持在 4

左右,再一次验证了 MIPDLM 模型中绝大部分 PDLs 能在 4 跳内搜索到所需资源。MIPDLM 模型较低的平均路径长度可减少网络延迟,减轻网络负载。

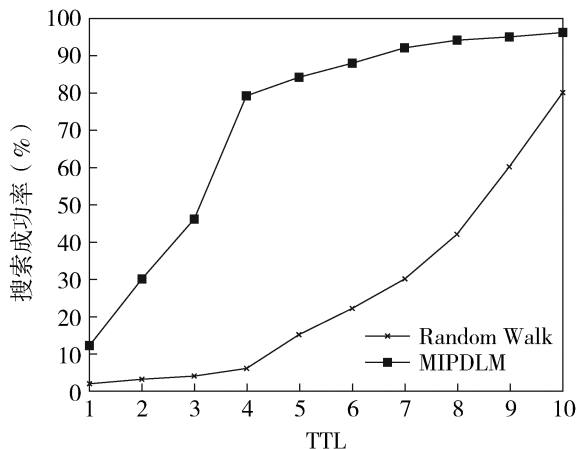


图 3 搜索成功率

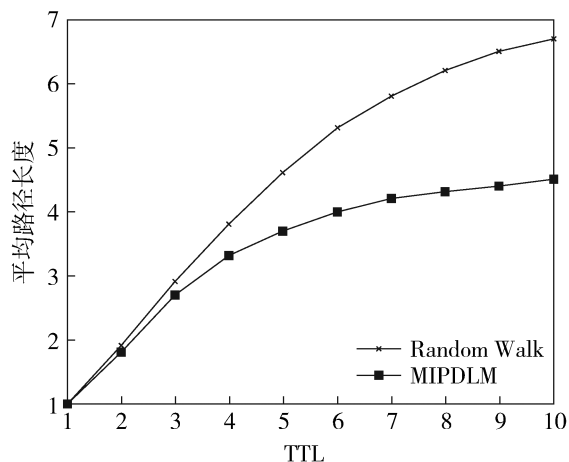


图 4 平均路径长度

## 4 结语

本文提出了一种基于多兴趣的 P2P 资源定位机制,个人数字图书馆可根据兴趣分簇探讨基于多兴趣的搜索机制。随着兴趣的动态变化,个人数字图书馆不断调整簇内邻居,系统始终将兴趣相似度较高的个人数字图书馆聚集在一起形成 P2P 兴趣覆盖网络。仿真实验显示, MIPDLM 模型中搜索消息不停地往兴趣相似度较高的个人数字图书馆转发,个人数字图书馆基本能在 4 跳内搜索到所需资源,可大大缩小搜索范围,提高搜索成功率,减少平均路径长度,减轻网络负载,提升个人数字图书馆的资源定位效率。本文为图书和档案共享系统的构建

提供的新思路,有助于建立高效的分布式资源定位机制和共享模型。

## 【参考文献】

- [1] 王文韬,谢阳群. 面向个人信息管理的一体化个人数字图书馆软件功能扩展研究[J]. 图书馆论坛,2012,32(3):94-97.
- [2] 陈光祚. 再论个人数字图书馆[J]. 图书馆论坛,2007,27(6):121-125.
- [3] 王小立. 百度“知道”知识传播对个人数字图书馆资源共享的启示:基于系统动力学方法[J]. 图书馆,2016(2):83-87.
- [4] 陈春艳. 试论个人数字图书馆的资源共享[J]. 情报杂志,2010,29(B06):302-303.
- [5] 张哲. P2P 网络下个人数字图书馆资源共享中个人隐私及知识产权保护方法[J]. 图书与情报,2009(6):34-40.
- [6] 王春梅,张银犬. 基于 P2P 技术的个人数字图书馆资源共享策略[J]. 情报杂志,2008,27(4):125-127.
- [7] Xu YA. P2P based personal digital library for community[C]//Sixth International Conference on Parallel and Distributed Computing Applications and Technologies (PDCAT'05), Dalian, 5-8 Dec. 2005 IEEE, 2005: 796-800.
- [8] 张银犬,朱庆华. 网络环境下个人数字图书馆信息检索策略

[J]. 中国图书馆学报,2007,33(3):56-59.

- [9] 王建勇,龚伏廷,李玉玲. 非结构化 P2P 网络中减少冗余的搜索策略[J]. 计算机工程与应用,2010,46(36):122-125.
- [10] 梅红岩. 基于节点局部信息与需求的非结构 P2P 网络搜索机制研究[D]. 北京:北京邮电大学,2014.
- [11] Sripanidkulchai K, Maggs B, Zhang H. Efficient content location using interest-based locality in peer-to-peer systems[C]//INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies. IEEE, 2003,3: 2166-2176.
- [12] 何可,吴晓军,张玉梅. 基于节点兴趣的非结构化 P2P 网络拓扑结构研究[J]. 计算机工程与应用,2016,52(9):102-107.
- [13] 陈香香,吴开贵,陈明. 基于兴趣域的对等网络动态搜索机制[J]. 计算机应用研究,2011,28(1):226-229.
- [14] Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks[J]. Nature, 1998,393(6684):440-442.
- [15] 马文明,孟祥武,张玉洁. 面向非结构化 P2P 网络的双向随机漫步搜索机制[J]. 软件学报,2012,23(4):894-911.

[收稿日期:2017-10-19]

[本文编辑:王颖]

## (上接第 29 页)

- [3] Top 100 overall collaborators[EB/OL]. [2017-06-26]. [http://www.natureindex.com/supplements/nature-index-2016-collaborations/tables/overall?utm\\_source=nature.com&utm\\_medium=Supplement-TOC&utm\\_campaign=Collaborations-2016](http://www.natureindex.com/supplements/nature-index-2016-collaborations/tables/overall?utm_source=nature.com&utm_medium=Supplement-TOC&utm_campaign=Collaborations-2016).
- [4] Greene M. The demise of the lone author[J]. Nature, 2007,450(7173):1165.
- [5] 杜建,张玠. 作者合作视角下的 h 指数计量方法:比较与归纳[J]. 图书情报工作,2011,55(24):52-55.
- [6] 中华人民共和国著作权法(2010 年修正)[EB/OL]. [http://www.sipo.gov.cn/zcfg/flfg/bq/fl/2015\\_09/t20150911\\_1174554.html](http://www.sipo.gov.cn/zcfg/flfg/bq/fl/2015_09/t20150911_1174554.html).
- [7] 娄策群. 社会科学评价的文献计量理论与方法[M]. 武汉:华中师范大学出版社,1999:148.
- [8] Cagan H, Sekercioglu. Quantifying coauthor contributions[J]. Science, 2008,322(5900):371.
- [9] Nils T. Hagen. Harmonic Allocation of Authorship Credit: Source-Level Correction of Bibliometric Bias Assures Accurate Publication and Citation Analysis[J]. Plos One, 2008,3(12):1-7.
- [10] Hodge SE, Greenberg DA, Challice CE. Publication credit[J]. Science,

1981,213(4511):950.

- [11] Nils T. Hagen. Harmonic coauthor credit: A parsimonious quantification of the byline hierarchy[J]. Journal of Informetrics. 2013, 7(4):784-791.
- [12] 贾贤,王霞,李忠富,等. 科技论文中等同贡献作者和共同通讯作者的署名问题[J]. 中国科技期刊研究. 2012,23(4):603-605.
- [13] GB/T 7713-1987, 科学技术报告、学位论文和学术论文的编写格式[S]. 北京:中国标准出版社,1987.
- [14] 秦小川,方红玲,刘雪立. 国内外生物医学期刊通讯作者标注情况的调查分析[J]. 中国科技期刊研究,2010,21(6):786-789.
- [15] Hagen, N. T. Harmonic publication and citation counting: Sharing authorship credit equitably—not equally, geometrically or arithmetically[J]. Scientometrics, 2010,84(3):785-793.
- [16] 李栎,张志强. 情报研究中核心著者的影响力评价方法研究[J]. 情报杂志,2010,29(10):80-83.

[收稿日期:2017-10-26]

[本文编辑:杨春华]