对等网信息检索的研究现状与展望*)

张 亮 邹福泰 马范援
(上海交通大学计算机科学与工程系 上海200030)

摘 要 随着对等网(P2P)研究的进一步深入以及P2P网络中Peer结点和共享文件的进一步增多,如何在非集中式的P2P网络中发现所需要的文件已经成为P2P从研究走向实用的关键所在。该文首先提出了P2P挖掘的概念,然后指出P2P信息检索作为P2P挖掘中的一部分,已经成为P2P研究的一个热点。接下来提出了P2P网络的路由、搜索、挖掘的框架模型,指明了该领域研究的框架。然后分层综述了P2P信息检索的进展状况,对各种检索方法做了深入分析。并指出了它们各自的优缺点和应用局限性,最后对今后的P2P信息检索领域的发展方向进行了展望。

关键词 对等网,信息检索,P2P路由,P2P挖掘,综述

A Survery and Prospects of Information Retrieval on the Peer-to-Peer Network

ZHANG Liang ZOU Fu-Tai MA Fan-Yuan
(Dept. of Computer Science and Technology, Shanghai Jiao Tong University, Shanghai 200030)

Abstract With the furthe development of the Peer-to-Peer research and the increase in the peer number and the documents published by peers in P2P network, how to retrieve the needed information in the P2P network is becoming more and more important. In the beginning of the paper the Pming(P2P Mining) concept has been proposed, and next this paper indicates that as a part of PMining, P2P Information Retreval(P2P IR) has been becoming a new research field. Next, a P2P routing, retrieval and mining model, which gives the research frame in P2P IR, is proposed. Next we review the development to P2P IR techniques. The characteristics of the P2P IR methods are analyzed; and the advantages, disadvantages and applied limitations are pointed out. Finally, the future direction in this field is discussed.

Keywords P2P information retrieval, P2P routing, P2P mining, Survey

1 引言

20世纪90年代,计算技术最引人注目的进展之一就是应用计算环境从集中走向分布,其中,Client/Server计算技术成为分布式计算的主流技术。随着技术的发展,基于Client/Server计算模式构建的商用计算体系结构表现出愈来愈大的局限性:它的网络优势局限于企业内部,难以突破企业之间的组织边界,企业与企业之间的信息交流受到很大制约,同时它还存在企业软件购置开销过大、网络安全性较差等诸多不足之处,这使得Client/Server计算模式已不能适应更高速度、更大地域范围的数据运算和处理。全球化、协作化、个性化决定了Browse/Server计算模式的出现。但这种模式要求在互联网上设置拥有强大处理能力和高带宽的高性能计算机,配合高档的服务器软件,再将大量的数据集中存放在上面,在服务器端集中处理数据,为互联网上其他PC进行服务。这种计算模式的缺点是没有有效地利用客户端的资源,包括处理器、存储器、文件内容等。目前,又出现了一种新的计算模式,就是P2P对等计算模式(Peer To Peer,简写为P2P)。P2P计算技术的出现就是希望能够充分利用互联网所蕴含的潜在计算资源。P2P技术的特征之一就是弱化服务器的作用,甚至取消服务器,任意主机既是服务器,同时又是客户机,即对等。对等网络实现了互联网的最终使用者到主动的服务提供者的转变。

回顾历史,计算模式的发展经历了从C/S到B/S再到P2P的一个过程;我们发现,在C/S模式下,数据通常存储在有固定结构的数据库中,在B/S模式下,其数据以半结构化的网页形式表现出来,而到了P2P模式下,数据的表现形式变成了无结构、多样化的文件;在C/S模式下,随着数据量的进一步增大,人们面临着数据爆炸但知识匮乏的尴尬境界,为了有效地解决这一问题,面向数据库的数据挖掘应运而生。在B/S模式下,随着Web页面数量的迅速增加,Internet成为了目前最大的知识库,对其进行模式和知识的获取已经成为必然趋势,因而面向Web的挖掘Web Mining应运而生。在P2P模式下,随着加入P2P网络的Peer结点的迅速增加以及发布文件的急剧增多,信息无序和信息过载即将成为一个越来越迫切需要解决的问题。在P2P网络中如何获取自己所需要的东西已经变得越来越困难,这时候,我们认为,针对于P2P文件的模式和知识的挖掘P2P Mining(简称为PMining)即将应运而生,并将成为下一个研究热点。(见图1)

简单地说,P2P挖掘就是使用数据挖技术自动地从P2P网络的Peer文件和服务中发现和提取信息和知识的技术。它是一个交叉的研究领域,涉及到传统的数据挖掘、文本、图像、声音、视频等方面的信息检索(Information Retrieval,简称IR)和信息抽取(Information Extraction,简称IE)、人工智能、模式识别特别是机器学习和自然语言处理(Nature Information Process,简称NLP)等领域。目前人们的P2P研究热点正转向P2P挖掘中的信息检索问题(P2P IR),即是如

*)本文受到上海市科委基础研究重点项目(02DJ14045)和国家自然科学基金重大国际(地区)合作项目(60221120145)资助。张 亮 博士研究生,目前研究方向为P2P信息检索,Web挖掘以及机器学习。邹福泰 博士研究生,目前研究方向为P2P信息检索。马范援 教授,博士生导师。

何在信息资源分布在各个独立的节点上的P2P网络中高效地索引、查找、检索出所需要的信息资源。P2P网络的信息检索按检索对象来分可分为文本信息检索、图像信息检索、声音信息检索、视频信息检索以及其他文件的信息检索等方面。本文中指的信息检索仅仅为P2P网络中的文本信息检索。P2P网络的信息检索问题作为P2P从研究到实用转变的关键，使得包括微软、SUN和HP等大公司和Berkeley、MIT、Stanford等国外著名科研机构在内的众多科研团体纷纷加入到P2P信息检索的研究行列中来，这充分说明P2P信息检索即将成为P2P研究领域中的一个活跃的研究热点。

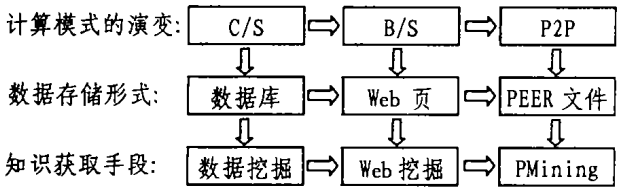

图1 计算模式-数据存储形式-信息挖掘的演化图

2 研究框架模型的提出

为了说明P2P信息检索领域的研究切入点，我们将P2P信息检索所牵涉到的研究重点分为三个层次，提出了P2P路由、搜索、挖掘框架模型(见图2)，该模型将路由、检索和挖掘在各个层次上进行了划分。该模型在overlay层之上又构架了两层，分别为文档表示层和用户模式层。overlay层上存储的内容为各结点的Peer ID，在该层次上的消息转发和结点定位我们称之为路由。而文档表示层上放置的是该Peer所拥有的共享文件发布的向量表示，在该层次上的获取相关发布文档的信息我们称之为检索，用户模式层表示的是用户/节点的模式、兴趣等信息，在该层次上为了提高检索效率和效果而进一步考虑用户模式我们称之为挖掘。按该模型来分，以往Peer-to-Peer网络的研究重点主要是overlay层上的改进路由效率问题，却很少从文档的表示、发布和查找角度以及从用户模式来提高信息检索的效率和效果。

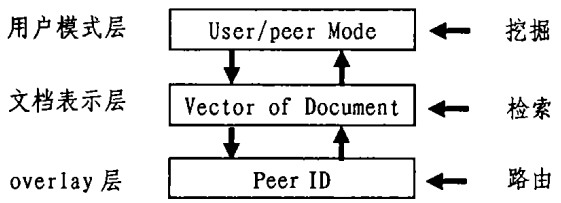

图2 P2P网络的路由、搜索、挖掘的框架模型

3 研究现状

我们在上节中提出了P2P路由、搜索、挖掘框架模型，详细说明了P2P网络中的信息检索问题可研究的三个切入点。现对三个层次的研究现状分别展开论述。

3.1 overlay层上的研究现状

P2P overlay层上要解决的问题是消息的路由问题。对于路由，传统网络通常通过集中的方式来解决，比如DNS技术，每一级服务器维护了相应的DNS表，用来在网络位置和物理位置建立起一一对应，客户端在需要定位物理位置时才向服务器发起请求，从服务器获得定位信息。P2P overlay层上最简单的一种实现方法即是采用类似DNS的集中式路由。但由于其存在单点故障、单点瓶颈等问题，故该领域的研究者多倾向于非集中式路由算法研究。在非集中式的路由算法中，关键问题是如何解决P2P非集中式网络中的Peer快速定位/路由问题。由于没有服务器/客户机这种主从关系，没有服务器端的集中式目录服务，为了快速地找到目标结点，如何对Peer的关系进行建模，达到尽量高的效率呢?在这里，计算机科学家们借用并发展了社会科学家们针对于社会中的小世界现象提出的小世界模型。

3.1.1 小世界模型 1967年，哈佛大学教授Milgram[5]做了这样一个试验，在美国随机选出160个人，让他们每人设法通过自己的朋友链将一封信传递到一个指定的人手中。最后有42封信到达了目标人手中，而且这些信的平均中间传递人数为5.5人。相比较2亿美国人口来说，这个数字显得非常低。这就是非常有名的"小世界"现象，又因为实验证明任何两个素不相识的人中间最多只隔着6个人，也就是说只用6个人就可以将他们联系在一起，因此又称为"六度分离"理论。这一点给P2P网络中Peer的迅速定位提供了重要的启示。

最早对小世界现象的建模的是Pool和Kochen模型，该模型中每个节点均匀从所有其它节点中随机地选择少量的节点连边。但是这个模型不能很好地解释社会中的社区问题。也即尽管我们两个人相互认识，但你认识的绝大部分人我却都不认识。这点不符合我们的社会常识。随后，Watts & Strogatz[6]1998年在Nature上提出了Watts & Strogatz模型，该模型认为相识边分两种，每个结点拥有大量的local边(近邻)，少量的long-range边(远亲)。远亲从远节点中均匀随机选择。正是这个模型的提出，引起了P2P应用研究的迅猛发展。Freenet就是主要依据该模型建立的。另外，在2000年，Jon Kleinberg[7]泛化Watts & Strogatz模型在Nature上发了一篇文章。该模型提出选择远亲时，概率应与离该节点的曼哈顿距离d有关，两点以概率d^{-r}连边。并给出了即使仅拥有距离为1的近邻和一个远亲，其算法的链长也不会超过$O(\log^2 n)$。正是因为这篇文章的发表，引发了人们对P2P研究高潮的到来，一系列非结构化、结构化算法才被提了出来。

3.1.2 基于小世界模型的路由 小世界模型的进一步完善给非集中式P2P网络中Peer的迅速定位提供了重要的理论依据，即尽管P2P网络中Peer结点的急剧增多但网络的消息转发次数却增长很慢或几乎保持不变。

Peer-to-Peer网络按其overlay层的构建方式可分为结构化P2P网络和非结构化P2P网络。其中结构化P2P网络是指在构建overlay网络时严格按照某种方式来对Peer间节点的相连，使得构建的overlay网络具有某种结构，因此我们称之为结构化网络。而非结构化P2P网络是指P2P节点之间建立邻居关系形成一个overlay网络，但形成的overlay拓扑结构没有严格定义并且数据的存放和overlay网络拓扑没有什么紧密的关系，这样的网络我们称之为非结构化网络。根据Peer-to-Peer网络的不同，其overlay层上路由可分为两种方式：一种是非结构化路由；另外一种是结构化路由。

在非结构化P2P网络中没有overlay层的结构信息可以利用，因此其路由方式主要有泛洪式，随机泛洪等盲目路由算法，但其优点是由于其拓扑没有严格的限制，路由算法可以利用已有的路由信息来对overlay层的扑拓结构进行修改，以提高其路由效率。

在结构化P2P网络中其overlay层的结构信息可以利用，因此针对于不同的结构拓扑，其路由方式也各不相同，其

中较常见的有 Chord 算法、CAN、Tapestry、Pastry 等路由算法。其中 Chord 算法[19]是由 MIT 提出的一种利用环行拓扑结构 overlay 层的分布式路由协议。每个结点对应一个 m 位长度的名字空间，并保存该节点后面$2^1, 2^2, \cdots, 2^m$个结点的指针信息，如某节点不存在，则指向存在的节点编号不小于(即等于或大于)该 ID 号的第一个节点上，这样每个节点仅需要 O(logN)大小的存储空间来保存路由层的拓扑信息(finger table)。当希望路由到某 Peer ID 点时，可在本节点的"finger table"中查找等于或小于但最接近 hash(Key)的节点，把消息转发到该节点。该节点递归执行以上步骤，直到最终查找成功。该算法的时间复杂度为 O(logN)。Content Addressable Network(CAN)[14]是由 Berkeley 提出的一种 N 维空间分割拓扑构造 overlay 层的非集中式路由算法。CAN 模型将节点映射到 N 维的坐标空间，根据服务器的密度和负载情况该坐标空间分成 N 维坐标空间区域，每个物理机器对应一个区域，并维护由哈希函数映射到该区域的数据和邻居信息。查找时，根据目标点坐标，从本结点开始沿着它的邻点进行查询，直到找到目的节点。在 d 维空间中，每个节点维护2d 个邻居位置信息，其空间复杂度为：2d。平均查找长度为：$(d/4)(n^{1/d})$。Pastry[12]是 Microsoft 和 Rice 大学共同发起的对等网络路由算法。在该系统中，每个节点有一个标识号 ID，并且维护一个 $\log_{2^b}N$ 行，2^b-1列的路由表，一个邻居集和一个叶子集。Pastry 的查找算法是根据"关键字"将信息路由到节点。查找时，首先在叶子集中查找是否存在的目的节点，如果关键字落入叶子集中的 ID 范围内，信息就被转发到具有在数值上最近 ID 的节点，否则，从路由表中选择一个比现有节点有更多共同前缀的节点号，把查询转交给这个节点以便下一步查询，如果仍然不可能，则在叶子集中寻找其 ID 具有相同前缀长度但数值是最接近关键字的节点。依此进行，直到找到目的节点。这种算法通常需要 $\text{Log}_{2^b}N$ 跳才能完成查找操作。Tapestry[20]是 Zhao，Kubiatówicz 和 Joseph 等提出的一种新型的 Overlay 层拓扑构造的定位和路由算法，这些结构化路由算法实际上都是首先自己提出一种 overlay 层的拓扑构建方案，然后利用该拓扑的结构信息，再来对路由算法进行构造的一种思路。这些方案的共同问题是脱离实际的网络拓扑以及将 Peer 看作是完全"对等"的，实际上各 Peer 间存在极大的异构性，如果能在路由构造上充分考虑这种异构性，并在实际运行过程中动态适应异构环境，将会提高 P2P 系统的性能。目前针对节点异构性的研究主要分为三个方面，一种是基于实际物理拓扑的路由[10,18]，其主要从 Peer 的地理位置分布和 Peer 的带宽分布来考虑。一种是基于主机性能的，主要考虑 Peer 的计算能力和稳定性分布来提高路由效率。还有一种是基于流量意识的路由，主要研究如何避免 hot spot 的问题。

3.2 文档表示层的研究现状

在文档表示层上的相关操作可分为两类：(1)共享文档的发布；(2)相关文档检索。

1)*共享文档的发布* 有以下三种方式：

·集中式发布，即 P2P 网络中的各个节点将共享文件(通常来说仅仅是文件名)注册在中央服务器上以便其它节点进行查询，如 Napster。

·复制式发布，即 P2P 网络中各个节点不将共享文件进行集中注册，而是自己维护自己的共享文件索引，但为了保证共享文件的可用性，可采用复制的方法，具体的复制策略有均匀复制、比例复制和平方根复制[9]。

·哈希式发布，即 P2P 网络中的各个节点将共享文件的文件内容分词后 Hash 发布所有词，其存在的问题是发布效率低，空间浪费严重，并且词间无主次之分。对这一点具体的改进思路有：从现有的 P2P 检索系统模型上考虑，选用更好的模型；采用分类/聚类后发布为了提高发布效率；考虑词在文档中的位置，加权发布；考虑检索词的分布。这些改进将在展望中详细讨论。

2)*相关文档检索* 我们基于上面的三种信息发布方式将信息检索方式分为集中式信息检索、泛洪式信息检索和 DHT 式信息检索。

·集中式信息检索：传统的信息检索和 Web 信息检索都是基于集中式的，而现在的 P2P 信息检索最简单的解决方案就是像 Napster[2]、点点能[3]那样建造一个中央服务器，里面保存了所有注册过的文件，所有查询工作由中央服务器完成；之后，各节点采用点对点的方式直接通讯。很明显，中央服务器具有单点失效，可扩展性不佳等缺点。

·泛洪式信息检索：为了解决集中式信息检索中存在的单点失效问题，人们提出了基于泛洪的信息检索方式，实际系统有 Gnutella[4]、FreeNet 等。在该类算法中，每个结点仅仅维护本身的内容索引，而不在其他结点上做本结点内容的索引。因此，为了找到存有相关内容的结点，采用的方案是进行泛洪。如在 Gnutella 中，一个节点要进行检索的话，它就向它的邻居节点广播消息(泛洪)，如果它的邻居结点中满足要求就返回它的结果集，如果没有，将该查询向该结点的邻居转发，查找的范围用 TTL 参数来。尽管这种方式是完全非集中式的，有效地避免了单点失效问题。但是它采用了广播消息加上 TTL 控制来进行查找，这种查找方式的检索范围仅在发起者为中心的某个特定半径以内，在超出一定范围后不能进一步扩展，因此查找并不能保证可靠性。其可扩展性同样不佳。Marius Portmann 等实验证明了 Gnutella 协议在系统规模达到几千个节点时将无法保证有效运转。

另外，考虑到造成系统不可扩展的主要原因是由于查询使用了泛洪式的消息传递方式，因而造成了广播风暴并引起结点的过载，Yang 和 Garcia-Molina 在文[13]中提出了迭代深入、直接广度优先搜索以及邻居内容的本地索引等机制来最小化广播过程中的带宽消耗。在文[15]中提到可利用 hint caches，gossip 和超级结点机制来提高搜索效率。Qin Lv 等[17]提出随机泛洪来减少广播流量以及采用分布数据复制策略来提高网络性能，但这并不能从根本上解决可扩展性的问题。

·DHT 式信息检索：为了进一步解决泛洪式信息检索中存在的可扩展性问题，人们提出了分布式哈希表的思想，在该类算法中，每个结点不仅维护本身的内容索引，而且维护其他结点上部分特定内容的索引，维护该内容的索引结点 ID 可通过 Hash 得到。因此，为了找到存有相关内容的结点 ID，仅对该关键字进行 Hash 即可。至于如何快速地定位到该结点，常见的算法有 Chord、CAN、Tapestry、Pastry 和 Plaxton 等。它们都是基于 DHT(Distributed Hash Table)技术，提供一种可扩展的分布 Peer ID 的查找机制。该查找具有确定性、可扩展性能好、容错性能高等优点。但不支持模糊查找[16]。之所以这种做法不适于模糊查找是因为在文档层上相似的文档向量映射到 Hash 层变得不相似了，深层次原因是由于文本表示空间到实际 Peer ID 空间的影射不是等距同构影射所造成的。

3.3 用户模式层的研究现状

该层次的基本思想是用户兴趣爱好相同的可划分到同一

个社区中，当用户需要同样的信息时，可以考虑首先在社区内查找。另外，可根据用户兴趣爱好来提供个性化信息检索服务，以便提高信息检索的效果。该层需要考虑的问题有：对个人偏好建模、如何引入用户的相关性反馈以及在P2P网络的用户模式层上建立/发现社区等。

4 进一步展望

结合上节提出的框架模型，可以看出目前P2P信息检索研究中仍存在着很多有待解决的问题。本文认为，在将来的研究中，以下几方面的问题可望成为该领域的主要研究内容：

(1)目前在P2P信息检索中几乎全部使用的都是布尔模型(Boolean Model)，尽管这一点并没有被明确声明。而布尔模型是基于集合论的一种简单匹配模型，其查询项(Query term)定义为包含该term的文档集合，Query为一个布尔表达式，查询项用布尔代数的运算符号来连接，其运算结果的文档集合作为检索结果，其主要缺点是无法在匹配结果集中进行相关性的排序，同时也无法区分词条在文档中所占的权重，并且漏检比较严重，因此总体来说，布尔模型是一种简单但是不够理想的检索模型。因此在文档表示上应采用更为贴切的模型。如采用基于词条的向量空模型、潜在语义模型或者基于语义的概念模型来表示。在向量空间模型(Vector Space Model)中，文档用加权的关键词向量(weighted keyword vector)来表示，相似度用两个向量的夹角余弦来计算，该模型优点是比较简单，易于计算，通常权值计算使用的是基于统计分析、直觉和试验的基础上的经验公式，典型的是TF/IDF的线性组合。但由于该模型term间相互独立的前提假设有些过于简化，容易造成误检(检索到不相关的文档，例如在一词多义情况下)和漏检(没有检索到相关的文档，例如在一义多词情况下)。潜在语义模型(Latent Semantic Index)的基本思想是将Term-Document矩阵进行SVD分解，将查询表达映射成缩减的文档向量空间中的向量，然后再进行相似度的计算。然而概念模型是采用树状或网状结构来表示概念的组织和分类，使用概念模型检索，就不再局限于词条本身，当用户输入一个查询词条时，不仅要找出与查询表达式匹配的结果，而且搜索引擎根据该词语概念与其他词语概念的内在关联，也要找出包含与查询表达式概念相同或相近的词语的文档。

(2)为了适应于全文检索，必须提高文档的发布效率，这方面可以采用分类/聚类后发布，先分为大类，再在其中聚小类。而不是每文档发布。这样可以大大减少发布关键词的数量，具有大大减少网络的索引空间，节省带宽等优点。另外，还可以考虑词在文档中的位置，如该关键词在标题中，进行加权发布等。同时考虑用户提交的检索词的分布来进一步提高发布效率。

(3)以往的研究重点都在P2P文件名搜索上，而且采用的是通常IR领域中的布尔模型(尽管没有明确提出)，决定了其不支持模糊查找。这个问题反映了在文档表示层上相似的文档向量映射到路由层变得不相似了。其深层次原因是由于文本表示空间到实际Peer ID空间的影射不是等距同构映射所造成的。解决方案有：1. 找到一个从文本表示空间到实际PEER ID空间的等距同构影射，Chunqiang在文[11]中提出了一种构造方法，但其缺点是该映射只能建立在CAN上，且两个映射空间的维数要匹配。2. 考虑在文档表示层建立文档间的相关性，将路由层的DHT和文档表示层的泛洪结合起来，该泛洪不同于路由层的泛洪，区别在于：1. 路由层的泛洪是以查询请求结点自身为中心的，现在的文档表示层的泛洪是以包含相似文档的某个结点为中心的；2. 路由层的泛洪中心仅有一个，现在的文档表示层的泛洪中心可有多个；3. 路由层的泛洪是全网络范围内的泛洪，现在的文档表示层的泛洪是相似结点内的泛洪。

(4)对用户模式层进行研究，将用户兴趣爱好相同的放入同一个社区中。当用户需要同样的信息时，可以考虑首先在同一社区内查找，若找不到，才到社区外查找。也即是说，考虑按用户/Peer的模式来进行社区划分，以提高信息检索的精度。可以用用户检索的关键词集合来构造用户访问模式。每个关键词集表示用户的某一兴趣点，可采用数据挖掘和机器学习等多种技术对用户发出的查询请求和用户已下载过的查询结果进行分析和训练，最终获得用户访问模式。

(5)如果能为P2P信息检索建立一个统一的理论模型的话，不仅可以为P2P检索技术的理论研究提供方便，还有利于促进P2P信息检索在应用领域中的发展。

结束语 本文首先提出了P2P挖掘的概念，然后指出P2P信息检索作为P2P挖掘中的一部分，已经成为P2P研究的从理论走向实际应用的关键。接下来根据提出的P2P网络的路由、搜索、挖掘的框架模型，分层对P2P信息检索的进展状况进行了详细的论述，在总结并评述它们的发展状况与优缺点的基础上，对该领域的发展方向进行了展望。

参 考 文 献

1 NS-2 Internet Website. URL: http://www.isi.edu/nsnam/ns/, 2003
2 The napster homepage. In: http://www.napster.com
3 The pptong homepage. In: http: www.ddtong.com
4 The gnutella homepage. In: http://www.gnutella.wego.com
5 Milgram S. The small world problem. Psychol. Today 2, 60-67 (1967)
6 Watts D J, Strogatz S H. Collective dynamics of small-world' networks. Nature, 1998, 393: 440～442
7 Kleinberg J. Navigation in a small world. Natuer, 2000, 406: 845
8 Yang B, Garcia-Molina H. Designing a super-peer network. In: 19th Intl. Conf. on Data Engineering, IEEE Computer Society. Bangalore, India, 2003
9 Qin Lv-Pei y Edith Cohen z Kai Li-Soctt Shenker. Search and Replication in Unstrctured Peer-to-Peer Networks. http://crystal.uta.edu/～kumar/cse6392/termpapers/Zhijun_paper.pdf
10 Yang B, Garcia-Molina H. Designing a super-peer network. In: 19th Intl. Conf. on Data Engineering, IEEE Computer Sciety. Bangalore, India, 2003
11 Reynolds P, Vahdat A. Efficient Peer-to-Peer Keyword Searching. http://issg.cs.duke.edu/search/search.pdf
12 Rowstron A, Druschel P. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In: Proc. IFIP/ACM Middleware 2001, Heidelberg, Germany, Nov. 2001
13 Yang B, Garcia-Molina H. Efficient search in peer-to-peer networks: [Technical Report 2001-47]. Stanford University, Oct. 2001
14 Ratnasamy S, Francis P, Handley M, Karp R, Shenker S. A scalable content-addressable network. In: Proc. ACM SIGCOMM. San Diego, CA, Aug. 2001. 161～172
15 Johansen H D, Johansen D. Improving object search using hints, gosip, and supernodes
16 Ratnasamy S, Shenker S, Stoica I. Routing Algorithms for DHTs: Some Open Question. In: Electronic Proc. for the 1st Intl. Workshop on Peer-to-Peer Systems (IPTPS'02)
17 Qin Lv, Pei Cao y Edith Conhen z Kai Li-Scott Shenker x. Search and Replication in Unstructured Peer-to-Peer Networks, 2002
18 Ratnasamy S, Handley M, Karp R, Shenker S. Topologically-Aware Overlay Construction and Server Selection, June 2002
19 Stoical I, et al. A scalable peer-to-peer lookup service for internet applications. Computer Communication Review v 31 n 4 Aug. 2001
20 Zhao B Y, Kubiatowicz J, Joseph A. Tapstry: An infrastructure for fault-tolerant wide-area location and routing: [Tech. Rep. UCB/CSD-01-1141]. University of California at Berkeley, Computer Science Department, 2001