

P2P 语义搜索研究进展

王志晓^{1,2} 张大陆¹ 王玉红¹ 王 东¹

(同济大学计算机科学与工程系 上海 201804)¹ (中国矿业大学计算机学院 徐州 221116)²

摘 要 P2P 语义搜索方法多种多样。首先,制定了一套综述方案,从语义描述方法、覆盖网络结构、查询扩展技术、资源聚类方式以及本体数量等 5 个角度对现有方法进行综述。然后,选择 10 种典型语义搜索方法进行分析比较,总结 P2P 语义搜索的主要特征。最后,指出下一步研究应探索新的资源组织与搜索机制,并着重关注 P2P 语义搜索性能评价、查询扩展技术完善、搜索结果排序等方面。

关键词 对等网,语义搜索,查询扩展,语义描述

中图分类号 TP393 **文献标识码** A

Survey of P2P Semantic Search

WANG Zhi-xiao^{1,2} ZHANG Da-lu¹ WANG Yu-hong¹ WANG Dong¹

(Department of Computer Science and Engineering, Tongji University, Shanghai 201804, China)¹

(College of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)²

Abstract A considerable amount of P2P semantic search approaches has been developed. First, the paper introduced a survey schema to review these semantic research approaches from five dimensions, i. e. semantic description method, semantic overlay network structure, query expansion technology, the way of resources clustering and ontology number. Then, the paper compared ten selected typical semantic search approaches by means of the schema. Based on this comparison, we identified and summarized outstanding features of P2P semantic search. Finally, the paper pointed out the necessity to explore new mechanism for resource organization and search, and to focus on a number of open problems, such as P2P semantic search performance evaluation, query expansion technologies improvement and search results ranking.

Keywords Peer-to-peer, Semantic search, Query expansion, Semantic description

Peer-to-Peer(P2P)是一种管理与共享网络资源的分布式计算模式,由于自身具有可扩展性、鲁棒性、容错性和自组织性等特点,受到了众多研究者的关注,在文件信息共享、协同计算和分布式存储等领域具有广泛应用。随着 P2P 应用的日益深入,人们可以从 P2P 网络上获得的资源越来越多。传统 P2P 网络基于关键字进行资源搜索,缺乏对语义的支持^[1]。于是,人们尝试将语义引入到 P2P 网络中来。出现了很多语义搜索方式,它们应用在不同的领域,其具体实现技术也不尽相同。

文献[2]从搜索(Search)、存储(Storage)、安全(Security)和应用(Applications)4 个方面对 P2P 搜索方法进行综述;文献[3]则从拓扑结构、数据存放策略、查询路由机制等角度分析基于 P2P 的 Web 搜索技术。文献[4]从基于文本的 P2P 搜索、基于内容的 P2P 搜索和语义搜索 3 方面对 P2P 内容搜索进展进行介绍。文献[2-4]只部分提及语义索引或搜索策略,没有专门针对 P2P 语义搜索方法进行综述。文献[5]从多个角度对语义搜索方法进行分类与综述,其主要针对搜索引擎领域,将 P2P 领域的语义搜索方法排除在外。目前,尚

未发现对 P2P 语义搜索方法进行全面综述的文献报道。

本文制定了一套对现有 P2P 语义搜索方法进行综述的详细方案,该方案涵盖了语义描述方法、覆盖网络结构、查询扩展技术、资源聚类方式以及本体数量等 5 个不同角度。本文还选择 10 种典型的 P2P 语义搜索方法进行分析、比较,总结 P2P 语义搜索的主要特征,发现现有方法的缺点与不足,明确下一步的研究方向与发展目标。在综述过程中,很难提及所有现存方法,但是要确保囊括所有的主流和代表性方法。

本文第 1 节给出 P2P 语义搜索方法综述方案;第 2 节利用制定的综述方案对典型语义搜索方法进行分析比较,总结 P2P 语义搜索方法的主要特征;第 3 节指出现有方法中存在的缺点与不足,明确下一步的研究方向与发展目标;最后是结论。

1 P2P 语义搜索方法综述方案

本节从语义描述方法、覆盖网络结构、查询扩展技术、资源聚类方式以及本体数量 5 个不同角度对现有 P2P 语义搜索方法进行综述。虽然也可以从其它角度进行综述,但我们

到稿日期:2009-08-28 返修日期:2009-11-12 本文受国家自然科学基金(90204010)资助。

王志晓(1979-),男,博士生,讲师,CCF 会员,主要研究方向为 P2P 语义搜索、语义场, E-mail: zhwxwang@cumt.edu.cn;张大陆(1949-),男,教授,博士生导师,主要研究方向为语义搜索、语义场、分布式系统。

认为这 5 个角度最能体现 P2P 语义搜索方法的主要特征。

1.1 语义描述方法

(1) VSM 或 LSI

向量空间模型(Vector Space Model, VSM)^[6]将资源表示成关键词及其权重的向量,每个关键词是向量的一维,这些关键词集合定义一个多维空间,每一个资源是多维空间中的一个向量。一旦将资源描述向量化,就可以采用现有的数学手段对这些向量进行处理。

VSM 模型只考虑关键词的信息,认为关键词条间相互独立,不考虑词与词在语义上的联系。实际上,关键词之间不可避免地存在着相互联系,如同义词、多义词。关键词间的这种相关性会影响系统性能,使搜索效果产生偏差。

潜在语义索引(Latent Semantic Indexing, LSI)^[7]根据关键词的共现信息探测词条之间隐含的或“潜在”的语义联系。LSI 通过奇异矩阵分解,将矩阵近似映射到一个潜在语义空间上,映射之后的奇异值向量能最大限度发现潜在语义联系,进而消除同义词、多义词的影响,提高资源描述的准确性。潜在语义索引利用统计分析和概率模型提取关键词间的语义联系,是一种浅层次的“统计”语义。

(2) Ontology

本体(Ontology)的概念起源于哲学领域,它对客观世界的事物进行分解,发现其基本组成部分,进而研究客观事物的抽象本质。本体是“概念模型的明确的规范说明”^[8]。“概念模型”指思考和描述现实中事物的抽象模型,通常限定在一定的领域,“明确”意味着抽象模型的概念和关系是确定的术语和定义。“规范”说明本体知识是可以被一个群体识别和接受的。

利用本体可以清晰准确地刻画资源的语义内涵,描述资源间的相互关系,在语义层次上解决资源描述问题,是领域内进行资源共享和互操作的语义基础,因此,本体在 P2P 中的应用越来越广泛,参见文献[9, 13, 29, 41, 47]等。

1.2 覆盖网络结构

(1) 传统 P2P 覆盖网络

在 P2P 语义搜索方法中,有些研究把重点放在语义描述方法、查询扩展技术等方面,没有过多地关注覆盖网络结构,采用的是传统 P2P 覆盖网络结构,参见文献[29, 36, 41]等。

(2) 语义覆盖网络

语义覆盖网络 SON^[9]的主要思想是将 P2P 节点按照语义信息聚类,以此提高资源搜索性能。在语义覆盖网络中,语义相近的节点组织在一起,形成 SON。每个节点统计自身资源的语义信息,根据统计结果判断自己归属的 SON,并加入该 SON。响应查询请求时,系统有目的地到符合该请求语义的 SON 中搜索,从而降低查询代价,提高查询结果的准确度。语义覆盖网络多采用层次结构。文献[10]提出了一种无监督的分布式 SON 构建方法 DESENT。

语义覆盖网络可以基于非结构化 P2P 网络,如 SemreX^[11],也可以基于结构化 P2P 网络,如 TTS-SON^[12];文献[13]提出了混合结构的层次化语义覆盖网络;文献[14, 15]则基于 Super Peer 结构来构建语义覆盖网络。

(3) 语义小世界

Watts 和 Strogatz 的研究^[16]表明,包括人类社交网络在内的许多网络都具有两个特性:较短的最短路径;较高的聚集

性。符合这种属性的网络称为“小世界网络”。小世界网络会导致“群居现象”,简单地讲就是具有类似兴趣的人通常会相互认识。为了提高 P2P 网络的查询性能,降低网络流量,可以将小世界理论应用到 P2P 语义搜索中。每个节点的邻居节点分类两类:短链(short links)和长链(long links)。短链是指与节点自身语义相近的节点,长链则从一类特殊节点中随机选取,这类节点拥有丰富的资源,并且资源集中于某一主题语义上。具有上述特征的 P2P 覆盖网络称为语义小世界^[17]。在 P2P 搜索过程中,可以从语义相近的短链节点上获取可能的搜索结果,也可以通过长链将搜索请求在 P2P 网络中快速扩散。文献[18, 19]是采用语义小世界结构的典型代表。文献[20]采用类似思想构建了一个语义社会覆盖网络 INGA。

(4) 语义链网络

语义链网络由网格及 P2P 领域专家诸葛海提出。在语义链网络中,节点与节点之间的语义关系称为语义链^[21]。语义链有因果链、蕴涵链、子类型链、相似链、实例链、顺序链和引用链等类型。语义链网络模型根据节点间的语义关系构建覆盖网络,是一个有向网络。当节点 P_i 加入 P2P 网络时,在网络中随机选择一个节点 P_j ,确定 P_i 和 P_j 的语义关系,把 P_j 作为 P_i 的后继节点。然后,根据语义链推理规则,得到 P_i 和网络中其它节点的语义关系。在响应用户查询请求时,根据节点间的语义关系及节点相似性,有目的地选择可能响应查询的后继节点。语义链网络可以基于非结构化 P2P,也可以基于结构化 P2P^[22, 23]。

1.3 查询扩展技术

Furnas 研究发现,两个人使用相同词汇描述同一事物的概率小于 20%^[24]。因此,必须对用户查询请求进行处理,以提高检索性能。查询扩展 QE(Query Expansion)^[25]是在初始查询的基础上加入与用户检索词相关联的新词,生成更准确的查询请求,弥补用户查询信息不足的缺陷,提高 P2P 网络中信息检索的查全率和查准率。

(1) 基于全局或局部分析的查询扩展

全局分析^[26]是对全部文档中的词或词组进行相关分析,计算每对词或词组间的关联程度,将词或词组按共同发生的频率先行聚类,其后根据词或词组的不同集合对查询进行扩展。局部分析^[27]利用初次检索得到的与原查询最相关的 top-k 篇文章作为扩展用词的来源,而非全部文档。局部分析法用到的技术有局部聚类、用户相关反馈(如用户日志、历史查询等)、局部上下文分析等。

(2) 基于本体的查询扩展

全局或局部分析在关键词匹配层次上进行查询扩展,难以充分表达用户查询的语义信息,不能从根本上消除用户查询意图与检索结果之间的语义偏差问题^[28]。文献[30]在全局分析法和局部分析法的基础上,通过构建语义关联树来进行查询扩展,但效果仍然不理想。

鉴于本体在语义描述方面的优势,研究者开始利用本体进行查询扩展。Voorhees 最早提出基于本体概念的查询扩展,并利用本体中的同义词和特定的子类进行扩展^[31]。文献[32]对基于本体的查询扩展进行了综述。文献[33]专门研究查询扩展中词语-概念相关度的计算。查询扩展常用的本体有两类:一类是通用词汇本体,如 WordNet, HowNet 等。这类本体在各个领域都可以广泛使用,缺乏专业领域的相关知

识;另一类是领域本体,该类本体应用范围有限,一般对领域内查询请求进行的扩展非常有效。

1.4 资源聚类方式

当前,多数研究都将语义相同或相似的资源聚集在一起,以提高语义搜索的查全率和查准率。资源聚类的具体实现有多种方式,这些方式可以归为两大类:间接聚类和直接聚类。

(1)间接聚类

间接聚类一般通过节点聚类达到资源聚类的目的。节点聚类的依据是节点语义描述。节点上一般都存有不同语义信息的资源,按语义信息统计节点上每类资源数量,可以将比例最高的资源类对应的语义信息作为节点的语义描述,也可以将节点的语义描述向量化,向量每一维对应某种语义的资源所占的比例。SON 就是典型的间接聚类。基于间接聚类的 P2P 语义搜索方法见文献[9,13,17,38,47]等。

(2)直接聚类

P2P 节点上的资源是多样化的,分布不均衡,且变化动态,所以很难准确、完整地通过节点聚类实现资源聚类。直接聚类避开了节点聚类中繁琐的资源语义统计过程,直接按照资源语义信息对资源进行聚类,具有相同或相似语义的资源的索引信息由特定节点保存,由这些节点响应针对该语义的查询请求。基于直接聚类的 P2P 语义搜索方法见文献[36,53]等。

1.5 本体数量

(1)单本体

单本体是指 P2P 环境中只有一个本体,所有节点采用相同的本体进行相关操作。一方面,单本体使不同节点对相同资源的理解与描述统一,可以很方便地在语义层次上实现资源的搜索与共享;另一方面,单本体降低应用系统的复杂性,系统简单,容易实现。面向社区环境的 P2P 应用一般都是单本体的。基于单本体的 P2P 语义搜索方法见文献[9,13,41]。

(2)多本体

一般情况下,不同领域的本体结构是不同的,即使同一领域,由于不同的需要、不同的侧重点,也会存在多个本体。因此,多本体更符合 P2P 网络的特征和实际情况。基于多本体的 P2P 语义搜索方法见文献[29,47]等。

为了在语义层次上实现多本体环境下的资源搜索与共享,必须解决多本体语义异构问题。一种方法是进行本体匹配^[34],发现不同本体元素间的语义关系,建立本体映射。另一种方法是把多个本体进行合并^[35],建立一个统一的单本体。

2 典型 P2P 语义搜索方法分析比较

本节选择 10 种具有代表性的 P2P 语义搜索方法进行分析比较。这 10 种方法是:(1) Hierarchical Summary Structure^[36](类似的方法还有文献[37,53]);(2)PMINER^[38](类似的方法还有文献[39,40]);(3)OI^[41](类似的方法还有文献[42-45]);(4)Semantic Overlay Network^[9](类似的方法还有文献[10,11,46]);(5)Two-tier SON^[13](类似的方法还有文献[12,14,15]);(6)H-Link^[47](类似的方法还有文献[48]);(7)SSW^[17](类似的方法还有文献[18,19,49]),虽然文献[50,51]也基于小世界理论构建覆盖网络,但并没有考虑节点及

源语义信息,不在本文讨论之列);(8)Semantic Link Networking^[52](类似的方法还有文献[21-23]);(9)SDQE^[53](类似的方法还有文献[54]);(10)Query Expansion and Interpretation^[29]。上述语义搜索方法的命名原则是:首选作者对其语义搜索方法命名,如果没有,则以该方法的核心技术与特征命名。分析与比较的结果如表 1 所列。对这 10 种方法的分析比较同样适用于上述提及的其它类似方法。

表 1 典型 P2P 语义搜索方法分析比较

项目名称	语义描述方法	覆盖网络结构	查询扩展技术	资源聚类方式	本体数量
Hierarchical Summary	VSM/LSI	传统 P2P 覆盖网络	无	直接聚类	无
PMINER	VSM/LSI	语义覆盖网络	无	间接聚类	无
OI	Ontology	传统 P2P 覆盖网络	无	无	单本体
Semantic Overlay Network	Ontology	语义覆盖网络	无	间接聚类	单本体
Two-tier SON	Ontology	层次化	无	间接聚类	单本体
H-Link	Ontology	语义覆盖网络	无	间接聚类	多本体
SSW	VSM/LSI	语义小世界	无	间接聚类	无
Semantic Link Networking	RSM	语义链网络	无	直接聚类	无
SDQE	VSM/LSI	传统 P2P 覆盖网络	全局分析	无	无
Query Expansion and Interpretation	Ontology	传统 P2P 覆盖网络	Ontology	无	多本体

从语义描述方法的角度来看,早期 P2P 语义搜索方法多采用 VSM/LSI 技术,如 Hierarchical Summary Structure, SDQE 等;VSM/LSI 技术描述的是一种浅层次的“统计”语义,影响了搜索性能的提高。随着本体技术的日渐成熟,本体在语义描述方面的优势体现出来,借助本体不但可以准确地刻画资源、节点的语义信息,还可以利用推理技术获取更为深层次的语义信息,很好地提高了 P2P 语义搜索的性能。因此,当前主流语义描述方法则采用本体技术,如 Two-tier SON, H-Link 等。

从覆盖网络结构的角度来看,部分研究把重点放在了语义描述方法、查询扩展技术等方面,采用的仍然是传统 P2P 覆盖网络结构,如 OI, Query Expansion and Interpretation 等。语义覆盖网络 SON 的概念提出后,研究者开始利用资源和节点的语义信息重构 P2P 覆盖网络,以提高语义搜索的性能,如 Two-tier SON, H-Link 等。除此之外,也有研究者提出了其它结构的 P2P 覆盖网络,如语义小世界、语义链网络。总体来说,语义覆盖网络 SON 是当前研究的主流与热点,出现了各种结构和性能的语义覆盖网络。

从查询扩展技术的角度来看,绝大多数 P2P 语义搜索方法没有对用户查询请求进行查询扩展。只有 SDQE, Query Expansion and Interpretation 等少部分研究将查询扩展技术应用到 P2P 语义搜索中,以弥补用户查询信息不足的缺陷,改善 P2P 网络中信息检索的性能。随着该项技术研究的深入,相信会有越来越多的 P2P 语义搜索方法采用它。

从资源聚类方式的角度来看,绝大多数 P2P 语义搜索方法采用间接聚类法,即通过节点聚类达到资源聚类的目的。P2P 节点上的资源是多样化的,分布不均衡,且变化动态,所以很难准确、完整地通过节点聚类实现资源聚类。Hierarchical

Summary Structure和 Semantic Link Networking 采用直接聚类法对资源索引进行聚类,避开了节点聚类中繁琐的资源语义统计过程,但是这些方法只是对局部资源进行直接聚类,目前还缺乏对全部资源进行直接聚类的方法。

从本体数量的角度来看,大部分项目研究单本体环境下的 P2P 语义搜索,只有少部分项目研究多本体环境下的 P2P 语义搜索,如 H-Link 采用 VSM/LSI 技术进行语义描述的 P2P 语义搜索方法不涉及本体。单本体是一种理想假设,多本体则更符合 P2P 网络的实际情况。

综上所述可以看出,当前 P2P 语义搜索方法多采用本体技术进行语义描述,并优化覆盖网络结构,利用节点和资源语义信息构建语义覆盖网络,资源聚类方式正在从间接聚类向直接聚类转变。越来越多的研究开始关注多本体环境下的 P2P 语义搜索,并逐步采用查询扩展技术来提高语义搜索性能。

3 下一步研究方向

3.1 探索新的资源组织与搜索机制

通过第 2 节的分析比较可以看出,P2P 语义搜索已有了一定的研究成果,但这些成果或多或少存在一些缺陷与不足,例如,当前主流语义覆盖网络 SON 存在诸如语义粒度划分^[13]、全局资源分布及网络拓扑^[10]等问题,另外,大多数 P2P 语义搜索方法所采用的间接聚类还存在聚类难的问题^[55]。由此可见,P2P 语义搜索领域还需创新性的研究,需要探索新的机制来解决复杂环境下基于语义的资源组织与搜索问题。

P2P 资源具有确定的语义信息,不同语义间是相互联系的,语义上越接近,联系越紧密,资源间呈现出一种按语义信息聚类的特性。如果能够采用一种数学模型描述这种相互联系,刻画这种聚类特性,就可以很好地解决基于语义的资源组织与搜索问题。场的概念最初用于描述物质粒子间的非接触相互作用,后来人们将其抽象为一个数学模型,用于描述对象间的非接触相互作用,揭示物理量在空间内的分布规律。目前,已有研究者用场模型描述数据空间中数据对象间的相互作用与聚类特性^[56]。将场理论引入 P2P 网络来解决资源组织与搜索问题,是一个可能的解决思路。

3.2 P2P 语义搜索性能评价

当前,P2P 语义搜索方法性能评价没有客观、统一的标准。下一步要建立完善的 P2P 语义搜索性能评价框架与机制,一方面可以评价现有方法,另一方面可以为设计新的语义搜索方法提供量化依据。P2P 语义搜索性能应从多个角度进行评价。从用户的角度看,主要考虑返回结果的查全率、查准率以及结果返回时间的长短等;从覆盖网络的角度看,主要考虑语义覆盖网络构建的代价、覆盖网络更新的复杂程度等;从资源索引的角度看,主要考虑索引在 P2P 网络中的负载均衡程度、资源索引更新与维护的成本等;从底层网络的角度看,则主要考虑 P2P 网络中产生的消息量。

3.3 查询扩展技术的完善

传统查询扩展主要采用全局分析或局部分析,难以充分表达用户查询的语义信息,不能从根本上消除用户查询意图与检索结果之间的语义偏差^[28]。鉴于本体在语义描述方面的优势,近年来,越来越多的研究开始利用本体进行查询扩展^[31,32]。

基于本体的查询扩展技术还有待完善,主要表现在:(1)

随着扩展层数的增加,扩展词数量会快速增长,可能导致出现查询偏移,层数越多查询偏移越严重。如何确定查询扩展层数是一个尚未解决的问题;(2)用户初始查询词最能反映用户需要,一般认为在使用过程中应对初始查询词赋较高的权重,对扩展词赋较低权重,至于权重如何具体设定没有很好的方法,通常依靠经验值给出;(3)在实际中,并非用户的每个查询词都对应在本体中,如果查询词不在本体中,如何选择合适的本体概念对其进行扩展是一个有待研究的问题。

3.4 搜索结果排序

大多数 P2P 语义搜索方法对查询结果不做处理,直接将其返回给用户。为了提高系统性能,在返回给用户前,应该对结果进行排序,确保语义最符合用户要求的资源排在最前面。当前,已有研究对这一问题进行了初步尝试,如文献^[55],该方法从词频的角度,而不是返回结果与用户请求语义相似的角度对结果进行排序。文献^[56]提出了语义排序算法,从语义信息熵的角度对结果进行排序,其定义的语义信息熵和词频仍然有很大关系。文献^[57,58]则是针对 Semantic Web 领域搜索引擎的结果排序算法。总之,搜索结果排序问题有了一些初步研究成果,但要应用到 P2P 语义搜索中,真正从语义角度对结果进行排序还需进一步研究。

结束语 本文从语义描述方法、覆盖网络结构、查询扩展技术、资源聚类方式以及本体数量等 5 个角度对现有 P2P 语义搜索方法进行综述,并选择 10 种典型语义搜索方法进行分析比较,总结 P2P 语义搜索的主要特征。最后,本文指出了当前研究的缺点与不足,明确了下一步研究应着重关注 P2P 语义搜索性能评价、查询扩展技术的完善、搜索结果排序等方面。

参考文献

- [1] Kiyohide N, Hiroyuki M, Tomonori A. Design and implementation of a semantic peer-to-peer network[C]//Proceedings of 7th IEEE International Conference on High Speed Networks and Multimedia Communications. 2004:961-972
- [2] Risson J, Moors T. Survey of research towards robust peer-to-peer networks: search methods[J]. Computer Networks, 2008, 50(3): 3485-3521
- [3] 方启明,杨广文,武永卫,等.基于 P2P 的 Web 搜索技术[J]. 软件学报, 2008, 50(3): 3485-3521
- [4] Merabti M, Liu Zhu, Yu H, et al. Advances in peer-to-peer content search[J]. Journal of Signal Processing Systems, 2009
- [5] Mangold C. A survey and classification of semantic search approaches[J]. International Journal of Metadata, Semantics and Ontologies, 2007, 2(1): 23-34
- [6] Wong S K M, Ziarko W, Raghavan V V, et al. On modeling of information retrieval concepts in vector spaces[J]. ACM Transaction on Database Systems (TODS), 1987, 12(2): 299-321
- [7] Papadimitriou C H, Tamaki H, Raghavan P, et al. Latent semantic indexing: a probabilistic analysis[C]// Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems. Seattle, Washington, United States, 1998: 159-168
- [8] Gruber T R. Toward principles for the design of ontologies used for knowledge sharing[J]. Computer Study, 1995, 43: 907-928

- [9] Crespo A, Molina H G. Semantic Overlay Networks for P2P Systems[C]//Proceedings of Third International Workshop on Agents and Peer-to-Peer Computing, New York City, USA, July 2004
- [10] Doukeridis C, Norvag K, Vazirgiannis M. DESENT: decentralized and distributed semantic overlay generation in P2P networks [J]. IEEE Journal on Selected Areas in Communications, Special Issue on Peer-to-Peer Communications and Applications, 2007, 25(1):25-34
- [11] 陈汉华, 金海, 宁小敏, 等. SemreX: 一种基于语义相似度的 P2P 覆盖网络[J]. 软件学报, 2006, 17(5):1170-1181
- [12] 于婧, 伊鹏, 刘强, 等. TSS-SON: 一种基于结构化 P2P 网络路由的两级语义覆盖网络结构[J]. 高技术通讯, 2007, 17(7):693-697
- [13] Gu Tao, Zhang Daqing, Hung Keng Pung. A two-tier semantic overlay network for P2P search[C]//Proceedings of the 13th IEEE International Conference on Parallel and Distributed Systems, 2007:1-8
- [14] 乔百友, 王国仁, 丁琳琳. TBSN: 一种基于分类层次的 P2P 网络[J]. 计算机研究与发展, 2008, 45(5):803-809
- [15] Loser A, Naumann F, Siberski W, et al. Semantic overlay clusters within super-peer networks[C]//Proceedings of the International Workshop on Databases, Information Systems and Peer-to-Peer Computing in Conjunction with the VLDB, 2003:33-47
- [16] Watts DJ, Strogatz S H. Collective dynamics of small-world networks[J]. Nature, 1998, 393(6684):440-442
- [17] Li Mei, Lee Wang-Chien, Sivasubramaniam A. SSW: a small-world-based overlay for peer-to-peer search[J]. IEEE Transactions on Parallel and Distributed Systems, 2008, 19(6):735-749
- [18] Jin Hai, Ning Xiaomin. Improving search in peer-to-peer literature sharing systems via semantic small world[C]//Proceedings of the 15th EUROMICRO International Conference on Parallel, Distributed and Network-based Processing, 2007:31-38
- [19] Jin Hai, Ning Xiaomin, Chen Hanhua. Efficient search for peer-to-peer information retrieval using semantic small world[C]//Proceedings of 15th World Wide Web Conference, Scotland, UK, May 2006:1003-1004
- [20] Löser A, Staab S, Tempich C. Semantic social overlay networks [J]. IEEE Journal on Selected Areas in Communications, 2007, 25(1):5-14
- [21] 刘洁. 对等网络环境下基于语义的异构数据管理模型[D]. 北京:中国科学院研究生院(计算技术研究所), 2005
- [22] Zhuge Hai, Liu Jie, Feng L, et al. Query routing in a peer-to-peer semantic link network[J]. Computational Intelligence, 2005, 21(2):197-216
- [23] Liu Jie, Zhuge Hai. A semantic-based P2P resource organization model R-chord[J]. The Journal of Systems and Software, 2006, 79(11):1619-1631
- [24] Furnas G W, Landauer T K, Gomez L M, et al. The vocabulary problem in human-system communication[J]. Communication of ACM, 1987, 30(11):964-971
- [25] Baeza-Yates R, Ribeiro-Neto B. Modern information retrieval [M]. New York: Addison-Wesley-Longman, 1999
- [26] Chang Y, Ounis I, Kim M. Query reformulation using automatically generated query concepts from a document space[J]. Information Processing and Management, 2006, 42(2):453-468
- [27] 张筹, 张霞, 刘积仁, 等. 混合 P2P 环境下有效的查询扩展及其搜索算法[J]. 软件学报, 2006, 17(4):782-793
- [28] Mitra M, Singhal A, Buckley C. Improving automatic query expansion[C]//Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1998:206-214
- [29] Ventresque A, Cazalens S, Lamarre P, et al. Dealing with P2P semantic heterogeneity through query expansion and interpretation[C]//Proceedings of International Workshop on Data Management in Peer-to-peer systems (DaMaP), co-location with ED-BT, Nantes, France, 2008:3-8
- [30] 桑艳艳, 刘培刚, 李勇. 基于语义计算的查询扩展优化研究[J]. 情报学报, 2007, 26(5):704-710
- [31] Voorhees E M. Query expansion using lexical-semantic relations [C]// Proceedings of ACM SIGIR, Dublin, Ireland, 1994:61-69
- [32] Bhogal J, Macfarlane A, Smith P. A review of ontology based query expansion[J]. Information Processing and Management, 2007, 43(4):866-886
- [33] 田萱, 杜小勇, 李海华. 语义查询扩展中词语-概念相关度的计算[J]. 软件学报, 2008, 19(8):2043-2053
- [34] Euzenat J, Shvaiko P. Ontology Matching[M]. New York Inc: Springer-Verlag, 2007
- [35] Kois K, Vouros G A, Stergiou K. Towards automatic merging of domain ontologies: the HCONE-merge approach[J]. Web Semantics, 2006, 4(1):60-79
- [36] Shen Hengtao, Shu Yanfeng, Yu Bei. Efficient semantic-based content search in P2P network [J]. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(7):813-826
- [37] Zhou Aoying, Ling Bo, Lu Zhiguo, et al. Efficient semantic search in peer-to-peer systems[C]//Proceedings of the Fourth International Conference on Web-Age Information Management, WAIM'03, LNCS 2762, 2003:278-289
- [38] Ferreira R A, Koyuturk M, Jagannathan S, et al. Semantic indexing in structured peer-to-peer networks[J]. Journal of Parallel and Distributed Computing, 2008, 68(1):64-77
- [39] Tang C, Xu Z, Dwarkadas S. Peer-to-peer information retrieval using self-organizing semantic overlay networks[C]//Proceedings of ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, Karlsruhe, Germany, 2003:175-186
- [40] Jin Hai, Ning Xiaomin, Chen Hanhua, et al. Efficient query routing for information retrieval in semantic overlays[C]//Proceedings of the 21st Annual ACM Symposium on Applied Computing, Dijon, France, 2006:1669-1673
- [41] Rostami H, Habibi J, Livani E. Semantic routing of search queries in P2P networks[J]. Journal of Parallel and Distributed Computing, 2008, 68(12):1590-1602
- [42] Haase P, Schnizler B, Broekstra J, et al. Bibster-A semantics-based bibliographic peer-to-peer system[J]. Journal of Web Semantics, 2005, 2(1):99-103
- [43] Kanellopoulos D N, Panagopoulos A A. Exploiting tourism des-

- tinations' knowledge in an RDF-based P2P network[J]. *Journal of Network and Computer Applications*, 2008, 31(2): 179-200
- [44] Nejdil W, Wolf B, Qu Changtao, et al. Edutella: a P2P networking infrastructure based on RDF[C]//*Proceedings of the 11th International World Wide Web Conference*. Honolulu Hawaii, USA, 2002: 604-615
- [45] Cai Min, Frank M. RDFPeers: A scalable distributed RDF repository based on a structured peer-to-peer network[C]//*Proceedings of the 13th International World Wide Web Conference*. New York City, USA, 2004: 650-657
- [46] Pagani E, Rossi G P, Pertoso E. ONRION-ontology-based query routing in overlay networks[J]. *Journal of Parallel and Distributed Computing*, 2009, 69(1): 28-38
- [47] Montanelli S, Castano S. Semantically routing queries in peer-based systems: the H-Link approach[J]. *Knowledge Engineering Review*, 2008, 23(1): 51-72
- [48] Jung J J, Zimmerman A, Euzenat J. Concept-based query transformation based on semantic centrality in semantic peer-to-peer environment[C]//*Proceedings of 9th Asia-Pacific Web Conference, Joint with the 8th International Conference on Web-Age Information Management*, LNCS 4505. 2007: 622-629
- [49] Huo Ying, Chen Zhigang. Searching scheme in P2P system based on semantic overlay network[J]. *Journal of Southeast University (English Edition)*, 2006, 22(3): 330-333
- [50] Zhuge Hai, Sun Xiaoping. A virtual ring method for building small-world structured P2P overlays[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2008, 20(12): 1712-1725
- [51] Hui K Y K, Lui J C S, Yau D K Y. Small-world overlay P2P networks; construction, management and handling of dynamic flash crowds[J]. *Computer Networks*, 2006, 50(15): 2727-2746
- [52] Zhuge Hai. Autonomous semantic link networking model for the knowledge grid[J]. *Concurrency and Computation: Practice and Experience*, 2007, 7(19): 1065-1085
- [53] Zhu Xing, Cao Hualiang, Yu Yong. SDQE: towards automatic semantic query optimization in P2P systems [J]. *Information Processing and Management*, 2006, 42(1): 222-236
- [54] Witschel H F, Bohme T. Evaluating profiling and query expansion methods for P2P information retrieval[C]//*Proceedings of the 2005 ACM Workshop on Information Retrieval in Peer-to-Peer Networks*. 2005: 1-8
- [55] Parreira J X, Michel S, Weikum G. p2pDating: Real life inspired semantic overlay networks for web search[J]. *Information Processing and Management*, 2007, 43(3): 643-664
- [56] 李德毅, 杜鹁. 不确定性人工智能[M]. 北京: 国防工业出版社, 2005: 187-217
- [57] 凌波, 周水庚, 周傲英. P2P 信息检索系统的查询结果排序与合并策略[J]. *计算机学报*, 2007, 30(3): 405-413
- [58] 郭炜, 何丕廉. 面向领域的语义排序与自适应热点算法[J]. *天津大学学报*, 2007, 40(7): 877-881
- [59] Ning Xiaomin, Jin Hai, Wu Hao. RSS: a framework enabling ranked search on the semantic web[J]. *Information Processing and Management*, 2008, 44(2): 893-909
- [60] Spiliopoulos V, Kotis K, Vouros G A. Semantic retrieval and ranking of semantic web documents using free-form queries[J]. *International Journal of Metadata, Semantics and Ontologies*, 2008, 3(2): 95-108

(上接第 20 页)

- [52] Cai M, Revesz P. Parametric R - Tree : An Index Structure for Moving Objects[C]//*Proc. of the Intl. Conf. on Management of Data (COMAD)*. 2000: 57-64
- [53] Prabhakar S, Xia Y, Kalashnikov D V, et al. Query Indexing and Velocity Constrained Indexing: Scalable Techniques for Continuous Queries on Moving Objects[J]. *IEEE Trans. on Computers*, 2002, 51(10): 1124-1140
- [54] Saltenis S, Jensen C S. Indexing of Moving Objects for Location-based Services[C]//*Proc. of the Intl. Conf. on Data Engineering (ICDE)*. 2002: 463-472
- [55] Procopiuc C M, Agarwal P K, Har-peled S. STAR-Tree: An Efficient Self-adjusting Index for Moving Objects[C]//*Proc. of the Workshop on Algorithm Engineering and Experimentats (ALENEX)*. 2002: 178-193
- [56] Tao Y, Papadias D J S. The TPR * -Tree: An Optimized Spatio-temporal Access Method for Predictive Queries[C]//*Proc. of the Intl. Conf. on Very Large Data Bases (VLDB)*. 2003: 790-801
- [57] Patel J M, Chen Y, Chakka V P. STRIPES: An Efficient Index for Predicted Trajectories[C]//*Proc. of the ACM Intl. Conf. on Management of Data (SIGMOD)*. 2004
- [58] Liu Z H, Liu X L, Ge J W, et al. Indexing Large Moving Objects from Past to Future with PCFI-Index[C]// *Advance in Data Management (COMAD)*. 2005: 131-137
- [59] 吕俊龙. 时空数据库中的移动对象索引技术研究[D]. 重庆: 重庆邮电大学, 2006
- [60] Pelanis M, Saltenis S, Jensen C S. Indexing the Past, Present, and Anticipated Future Positions of Moving Objects[J]. *ACM Trans. on Database System*, 2006, 31(1): 255-298
- [61] 刘晓军. 一种高效的移动对象位置索引机制的研究[D]. 济南: 山东大学, 2007
- [62] Tao Y, Papadias D, Zhang J. Cost Models for Overlapping and Multi-version B-trees[C]//*Proc. of the 18th Intl. Conf. on Data Engineering (ICDE)*. 2002: 191-200
- [63] Choi Y J, Chun C W. Selectivity Estimation for Spatio-Temporal Queries to Moving Objects[C]//*Proc. of the ACM Intl. Conf. on Management of Data (SIGMOD)*. 2002: 440-451
- [64] Theodoridis Y, Silva J R, Nascimento M A. On the Generation of Spatio-temporal Datasets[C]// *the 6th Intl. Symp. on Large Spatial Databases (SSD)*. 1999: 147-164
- [65] Theodoridis Y, Sellis T, Papadopoulos A N, et al. Specifications for Efficient Indexing in Spatiotemporal Databases[C]//*Proc. of the Intl. Conf. on Scientific and Statistical Database Management (SSDBM)*. 1998: 123-132