

网络空间大搜索研究范畴与发展趋势

方滨兴^{1,2}, 贾焰², 李爱平², 殷丽华³

(1. 北京邮电大学 计算机学院, 北京 100876; 2. 国防科学技术大学 计算机学院, 湖南 长沙 410073; 3. 中国科学院 信息工程研究所, 北京 100093)

摘要: 随着网络空间的拓展、网络应用模式的发展及大数据时代的到来, 面向网络空间的下一代搜索引擎——“大搜索”已具有迫切的需求。阐述了网络空间大搜索的内涵及其特点, 提出了大搜索的研究范畴, 包括泛在网络空间信息获取与发掘、知识仓库构建和管理、用户搜索意图准确理解与表示、用户意图高效匹配和推演、大搜索安全可信与隐私保护等方面的内容, 并针对上述问题, 指出了具有5S特性的网络空间大搜索技术的发展趋势。

关键词: 网络空间大搜索; 意图理解; 意图匹配; 智慧解答

中图分类号: TP393

文献标识码: A

Research progress and trend of cyberspace big search

FANG Bin-xing^{1,2}, JIA Yan², LI Ai-ping², YIN Li-hua³

(1. School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. School of Computer Science, National University of Defense Technology, Changsha 410073, China;

3. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China)

Abstract: With the expansion of cyberspace, the development of Web application and the arrival of the era of big data, searching for the cyberspace has become an urgent requirement. The concept “cyberspace big search” was proposed and its basic features was discussed. Based on these features, the architecture for the big search was proposed, including five parts: acquiring and mining information from the cyberspace, constructing and managing knowledge wares, understanding and representing the user’s search intention, matching and reasoning the search intention, security and privacy in big search. The research progress and trend of cyberspace big search are summarized.

Key words: cyberspace big search; intention understanding; solution matching; intelligent solution

1 引言

搜索引擎是指从互联网搜集信息, 经过一定整理以后, 提供给用户进行查询的系统。搜索引擎填补了人们和信息空间之间的信息鸿沟, 在促进互联网飞速发展、加快互联网应用普及等方面, 起到了非常重要的作用。然而, 随着网络空间逐渐由互联网扩展到由互联网、物联网、传感网、社交网络、工业控制系统等信息系统所构成的泛在网络空间发展, 网络应用模式也从 Web 1.0 发展到了 Web 3.0^[1], 随着大数据时代的到来^[2], 搜索内容已由单

纯信息向多维度拓展, 传统的搜索引擎已经不能满足需求, 新时代的搜索——“网络空间大搜索”应运而生。

网络空间大搜索是具有“智慧”的搜索, 在网络空间技术及应用蓬勃发展的今天, 将在未来对我国政治、经济、技术、生活等各个方面产生重要的影响, 具有极其重要的战略意义, 具体包括如下。

1) 影响和推动经济发展。搜索引擎将以几何基数增长的形式带动相关行业经济发展。首先, 用户数增长是基础, 2014年中国移动搜索用户已达到4.5亿, 移动搜索市场规模比2010年上涨32倍。其次, 广告

收稿日期: 2015-10-12; 修回日期: 2015-12-15

基金项目: 国家重点基础研究发展计划(“973”计划)基金资助项目(2013CB329601, 2013CB329604, 2013CB329606); 国家自然科学基金资助项目(61472433, 61502517)

Foundation Items: The National Basic Research and Program of China (973 Program)(2013CB329601, 2013CB329604, 2013CB329606); The National Natural Science Foundation of China (61472433, 61502517)

增长是收入来源, 2013 年 11 月, Google 在线广告推荐系统收入已经超过全美报纸、杂志行业广告收入的总和。Gartner 预计全球移动广告支出有望由 2014 年的 180 亿美元增加至 2017 年的 419 亿美元。2014 年中国搜索引擎市场迎来较为高速的发展, 市场规模达到 587.2 亿元, 到 2017 年将翻一番。

2) 支撑和推动 IT 技术进步。网络空间大搜索将搜索对象由单纯的互联网信息扩展到了人、物体和信息组成的三维信息空间。三维信息空间具有明显的 5V 大数据特征(规模大、类型多、速度快、不确定性、价值性)。据 IDTechEx 预测, 未来 10 年全球的智能终端节点数量将达到 10 万亿。2013 年 5 月中国工业和信息化部发布的《物联网标识白皮书》表明, 未来我国的物联网终端数量将达到 100 亿~1 000 亿。在 2011 年全球产生互联网数据量已经达到 1.8 ZB, 种类包括文本、视频、音频、图片等, 中国联通上网记录每秒 83 万条, 这一切都表明大数据无论是规模、种类还是产生速度, 都已产生了革命性变化。大搜索将以桥梁连接作用促进物联网、移动互联网、大数据、云计算四大 IT 技术的发展。正如搜索引擎与网络技术及 Web 应用的相辅相成的关系一样, 大搜索的发展也必将与四大 IT 技术相互促进, 推动信息产业的繁荣。

3) 提升人民的生活水平。大搜索将在环保、医疗、教育、交通等方面的应用, 将有效提高政府、企业、机构及个人的决策能力, 提供快捷、全面、准确的决策依据, 大大提高决策能力, 提升决策透明度。大搜索以发掘大数据深层信息的“深挖”能力, 连接各行业全方位信息的“贯通”能力, 为用户提供问题解答的“智慧”能力, 有效提升人民的生活水平。

4) 为国家安全提供情报。搜索引擎可以通过技术手段操控人们获得信息的范围, 谁掌握了搜索引擎, 谁就拥有了为人们提供信息甚至答案的权利, 由此产生的政治、经济和社会驱动力日益受到各国重视。搜索引擎通过对用户的搜索问题进行战略性统计和计算, 这将对国家、社会和商业具有重要意义。

网络空间大搜索面向泛在网络空间的人、物体和信息, 在正确理解用户意图基础上, 基于从网络空间大数据获取的知识, 从信息、时间、空间的角度给出满足用户需求的智慧解答。其中, 用户意图理解是在用户多模态输入及消除歧义的基础上, 结

合用户的上下文和语义知识, 迅速、准确地理解和定位用户的真实意图, 从而缩短用户与搜索引擎的交互流程, 达到提高用户感受、获得精准搜索结果的目的。智慧解答是在综合利用大数据价值和知识体系基础上, 经过匹配、推理、计算以及众包等技术和方法, 形成若干个智慧的综合解决方案, 给出真正满足用户需求的可供选择的解决方案体系, 而不仅仅是基于简单匹配的存在性信息搜索结果。

2 相关工作

已有搜索引擎按其工作方式主要可分为 3 种, 分别是全文搜索引擎、垂直搜索引擎^[3]和元搜索引擎^[4]。其中, 全文搜索引擎是从互联网上提取各网站的信息创建数据库, 并将网页按相关度排序, 用户输入关键词, 搜索引擎将匹配的网页链接返回给用户^[5], 如 Google、Baidu 和 Bing 等; 垂直搜索专注于特定的搜索领域和搜索需求, 是搜索引擎的细分和延伸, 可划分为人搜索、物体搜索、服务搜索和领域搜索等^[6], 如大众点评网、同城网和房产搜索等^[7,8]; 元搜索引擎是同时在其他多个引擎上进行搜索, 并将结果返回给用户, 如 MetaCrawler 等^[9]。

一般地, 大搜索至少包含知识仓库的构建和管理、用户搜索意图的准确理解、表示与匹配、安全与隐私保护, 因此, 下面从这些方面来展示国内外现状。

2.1 知识仓库的构建和管理

知识仓库是知识管理与知识服务的基础, 目前知识仓库的构建和管理主要围绕知识图谱(knowledge graph)展开, 从而增加搜索的智能性, 提高用户体验。知识图谱的数据源主要包括 2 类, 一类是 Web 页面^[7~9], 一类是在线百科^[10~12]。在知识仓库的建立和管理中, 资源的有效表示与查询至关重要。RDF 作为 Web 资源的表示标准, 许多知识图谱都选择 RDF (或类 RDF) 来表示知识。早期 RDF 查询多先利用关系数据库中的关系表存储 RDF 数据, 而后期将 RDF 查询转换为 SQL 查询^[13,14]的焦点在于进一步提升 RDF 查询性能, 其相关技术包括垂直分片^[15]、额外索引^[16]等。目前大多知识均以简单 RDF 形式存储, 这使多数图查询模型及技术可应用到知识库查询处理中, 主要包括: 可达性查询^[17]、最短路径或距离查询^[18,19]、图匹配查询^[20,21]、序列查询及关键字查询^[23]。

已有的知识图谱技术偏重已知的实体, 对不断

涌现的新兴实体及其关联，尤其是事件性及隐性实体的关联，尚没有完善的解决方法。

2.2 用户搜索意图的理解、表示与匹配

只有充分理解用户真正的检索意图，才有可能快速准确地找到查询结果，为用户提供准确的检索服务。因此，对搜索意图进行建模是信息检索领域的研究重点之一。最直接方法是基于用户对搜索结果的点击行为进行搜索意图建模。如 Li 等^[24,25]提出一种半监督学习方法，利用查询词与搜索结果之间的点击二分图及部分已标注信息生成分类器，达到推断查询词对应的搜索意图的目的。另外，Etzioni 等提出了基于规则模板抽取实体/概念之间的关系来描述和理解搜索意图。Madhu 等^[26]利用语义网工具和技术提供分层模块的方法解决搜索引擎对语义内容的理解。

在基于用户点击行为的建模方法的基础上，引入时序相关性分析法分析用户一段时间内的搜索日志可以提高建模的准确性^[27~30]。Chilton 等^[27]通过分析大量的用户搜索及点击记录，发现用户重复的搜索行为与搜索关键词的价值之间的关联关系，进而利用这一关系判断用户之后的关键词输入行为所对应的准确搜索意图。He 等^[28]基于持续的部分可观测马尔可夫模型(POMD, partially observable Markov model with duration)分析用户搜索日志中包含的空间与时间信息，获取用户的一些隐性行为特征，并通过这些特征构建用户搜索意图模型。Agichtein^[29]则根据对用户的搜索记录的分析，提出一种对用户较长一段时间内可能发生的行为的预测方法，从而达到为用户提供准确检索结果的目的。

此外，研究者提出了基于语义相关性的意图建模方法。Sadikov 等^[31]结合用户查询后点击的文件和查询关键词等所包含的语义信息分析用户的查询意图。Jethava 等^[32]分析用户查询关键词除自身语义外所包含的多维信息，建立从粗糙到精细的树状结构，根据查询词的聚类结果进行意图建模。在构造查询词的树状结构时，这些算法都需要利用自然语言处理的方法对查询词之间的关系进行分析^[33]。Li 等^[34]同时结合搜索意图在时间上和语义上的相关性分析，提出了基于高维 Hawkes 点过程模型的用户搜索意图识别和分类模型，统一了上述建模方法。

用户意图匹配的研究工作主要包括文本模型和图模型。基于文本模型的意图匹配通过将以关键词查询检索的方式来把用户的意图进行语义转换

和目标文档的匹配，并获取相关度排序。基于图模型的意图匹配通过图搜索来实现搜索意图与搜索空间中目标项的查找和匹配，主要包括关键词图搜索技术、子图匹配技术和近似图匹配技术等^[35]。

虽然研究者在用户搜索意图的理解、表示与匹配方面已经取得了不少成果，但是现有的搜索意图理解多是经验式的，仅是分析性的而非预测性的，目前仍缺少统一的理论和算法框架对搜索意图的建模进行指导。此外，现有的基于搜索意图的交互式搜索方法往往是被动探索用户意图而非主动地启发用户更明确表达其搜索意图。另外，现有的研究几乎忽略了用户在与搜索引擎交互过程中表现出来的间接行为（如头部运动），研究这些间接行为将成为搜索意图建模的发展趋势。

2.3 大搜索安全与隐私保护

大搜索的数据可能来源于物联网、工控网、移动互联网，甚至是与国家安全相关的基础设施。相应地，攻击者可能利用大搜索系统获得个人隐私、商业甚至国家机密。因此，面向大搜索的安全与隐私保护至关重要。访问控制作为保障安全的关键技术之一，在搜索领域有广泛的应用。在大搜索中，数据与搜索者具有高度动态性、海量性等特征，这使适用于封闭环境的传统访问控制不再适合于大搜索。基于属性的访问控制将主体和客体的属性作为基本的决策要素，属性是主体和客体内在固有的，不需要手工分配，同时访问控制是多对多的方式，管理上相对简单。这些优点使基于属性的访问控制在大搜索中具有广阔的应用前景。

基于属性的访问控制中用户权限仅与其属性相关，具有较强的匿名性，为恶意用户滥用其所拥有的权限带来了方便，恶意攻击者可能利用大搜索作恶，因此必须研究基于属性的身份认证、追踪、权限更新与撤销机制。结合数字签名思想，文献[36,37]提出了基于属性的签名机制，该机制中签名者可声称其签名对应某一组特定的属性或某种特定访问控制结构，验证者可验证签名是否由相应的属性或访问结构拥有者所签署。根据追踪方法输入的不同，用户可追踪方法可分为白盒追踪(white-box traceability)^[38]及黑盒追踪^[39](black-box traceability)机制。在权限管理方面，Ostrovsky 等^[40]首先提出了基于 CP-ABE 的直接撤销机制，但这种方式增加了密文和用户私钥的大小。为了减小授权机构的负担同时实现细粒度的访问权限管理，研究者提

出了间接管理的方法^[41,42]。

在隐私保护方面,文献[43]将密文检索理论应用到物联网隐私保护中,用户可以在不泄漏位置信息的情况下检索服务器上的任意数据项。文献[44]提出一种基于可查询加密的隐私保护框架,在数据加密的情况下实现好友之间签到位置的查询。文献[45]提出了基于差分隐私的查询处理技术。文献[46,47]将分权机制引入到搜索代理平台,在搜索过程中搜索平台无法准确获知用户的身份信息,进而保护用户的搜索模式。

虽然研究者在面向搜索的安全与隐私保护方面取得了不少成果,但多数工作围绕着以访问控制等为核心的技术展开。大搜索模式具有扁平性、搜索用户的开放性与海量性、节点动态性等特征,在该模式下如何实现搜索权限的实时撤销与更新、如何在数据加密数据中实现搜索、如何有效分割数据、如何添加噪音以防止信息被搜索者推演获取有待进一步研究与完善。

3 网络空间大搜索特征

与传统的搜索引擎相比,网络空间大搜索具有以下“5S”特点。

1) 泛网获取(sourcing from the cyber)。根据给定的目标和任务获取数据,其数据空间是涵盖了人、物、信息的泛在网络空间,获取的数据类型包括人、物、事件、时间、空间等各类信息,并进行有效组织、存储和管理,为智慧解答奠定基础。

2) 意图感知(sensing the context)是对用户搜索意图的精确化理解。为此,大搜索结合用户请求的上下文、时空特性、场景感知、动作情感等方式,支持在语义级别上对用户搜索意图进行理解,并以恰当的方式进行表示,提交给搜索引擎,意图感知是大搜索的基础。

3) 多源综合(synthesis from multiple channels)是基于多模态数据(如文本、位置、传感器、交通、图片、音视频等数据)进行多源关联推理,并给出的多维度、多属性、多模态智慧解答。

4) 安全可信(security privacy and trust)要求搜索结果的可信性和用户的隐私保护。不同于传统搜索引擎给出的存在性搜索,大搜索需要对数据进行挖掘、分析和加工,在此基础上给出综合的解答,因此其结果的可信性至关重要;搜索中的关联分析可能挖掘出用户的隐私,因此需要进行隐私保护。

5) 智慧解答(intelligent solutions)。传统搜索引擎的搜索结果只包含用户输入关键词的匹配网页,而网络空间大搜索引擎返回的答案是经过理解和推理综合的解答。具体过程根据用户的搜索意图,基于知识仓库对关联的知识进行求解,通过推理、统计、众包等多种推理演算方法形成若干个智慧综合的解决方案,并将之以合适的方式提交给用户。

4 网络空间大搜索研究范畴

网络空间大搜索引擎填补人与物理世界、信息世界的断层,其体系结构如图 1 所示,其研究范畴主要包括 5 个部分。

1) 泛在网络空间信息获取与发掘方法。以一定的策略和方法,面向给定任务目标在网络空间中采集、获取和推演相关数据和信息,主要技术包括如下。

①面向目标任务的多来源、多模态数据获取方法:面向各种应用领域和不同数据模态的目标任务表示、匹配及获取技术;面向实时数据流的目标信息采集技术;目标驱动的异构、异质数据的协同采集技术;巨规模采集任务并行计算和管理平台技术;目标采集数据的完整性和精确性评估模型等。

②面向目标任务的关联数据发掘方法:数据关联推演知识的表示、管理及基于推演的间接数据获取方法;基于上下文的多模态数据关联挖掘方法;场景、时空感知的关联数据挖掘分析方法;基于众包、标注等方法的关联数据挖掘方法等。

③巨规模、多模态实时数据流的清洗方法:基于滑动窗口数据摘要及优先队列的重复数据删除技术;基于编辑距离算法的异构相似数据匹配技术;基于情景语义描述模型的噪音数据清洗技术等。

④泛在网络空间数据融合与冲突消解方法:基于数据依赖关系图的多模态异构数据的融合计算模型;情境驱动的多层次融合和情景语义描述模型;基于本体论的多层次(数据级、特征级、决策级)数据冲突消解方法等。

2) 知识仓库的构建和管理模型。在给出泛在空间巨规模实体关系统一建模知识表示的基础上,再对知识从不同维度、不同层次等方面进行聚合、组织和关联,并维护其最新状态,提供高效的查询、匹配和推演等操作,主要技术包括如下。

①巨规模实体关系的表示模型和方法:基于超图的统一实体关系表示模型;实体间巨复杂关联关系及其演化的表示方法;实体多维属性的及其时空

向情景感知的深度融合隐私保护机制,研究面向搜索的高效隐私保护理论;研究设计能够抵御关联分析的隐私保护策略;研究隐私保护方案的动态调整机制,实现对海量用户的高并发隐私保护方案。

从上面分析可以看出,网络空间大搜索引擎的研究范畴与传统搜索引擎存在重大差别,体现在如下几个方面。第一,是关联信息发掘,即对泛在网络空间中的数据获取和信息发掘,包括互联网、物联网、社交网络、医疗健康、视频监控、地理信息等空间中的各类数据。第二,是知识仓库构建和管理,即面向泛在网络空间的海量对象及关系进行建模,该模型支持语义级、巨规模实体、关系、及其演化的表示,并形成知识立方。第三,是搜索意图理解,即结合用户的上下文和语义知识等方法,迅速、准确地理解用户的真实意图。第四,是知识推演,经过匹配、推理、计算乃至众包等技术和方法,形成若干个满足用户真正意图的智慧综合的解决方案。第五,是安全隐私保护,即保证用户搜索的全程安全,使数据源可信、搜索过程可控、搜索结果可过滤。

由于大搜索从泛在网络空间中的数据获取、支持语义知识仓库构建和管理、能准确理解搜索意图、在保障安全与隐私的条件下形成满足用户真正意图的智慧综合解决方案。因此,相对于传统的互联网搜索,网络空间大搜索引擎将对用户体验与应用产生革命性的提升和改善。

5 网络空间大搜索技术发展趋势

在泛在网络空间大搜索引擎的基础理论发展趋势方面,尚需要针对大搜索研究范畴中的5个部分进行突破,具体包括:1)针对数据间显式或隐式形成的巨大实体关联网络,构建面向巨规模实体与数据搜索的知识仓库,设计与之相配的管理方法;2)结合用户上下文与环境等信息,实现对用户搜索意图的准确理解;3)解决海量异构数据的定向获取问题,并实现基于推演等的间接知识发现;4)基于子图匹配、计算统计、规则推理、众包等技术解决大规模/不完备的知识仓库与用户意图的实时匹配,有效地从大量数据中搜索定位目标实体和目标关系,实现秒级的知识匹配、推理和统计;5)大搜索中的安全与隐私保护问题,解决开放数据的源可信问题、搜索过程的可控问题、用户的隐私保护问题以及暴力色情等有害信息的过滤问题。

在网络空间大搜索的应用试验床发展趋势方面,主要是应对物联网、社交网络和视音频等领域需求,构建大搜索试验床,具体包括:1)安全物联网搜索试验床,覆盖多类传感器、摄像头、SCADA网络、位置服务等多种应用,研发基于物理实体发现的多维时空高动态索引系统,建立支持实时搜索的跨地域分布式物理实体与信息的安全搜索试验环境;2)社交网络搜索试验床,覆盖微博、博客、人人等多通道,集成Twitter、新浪微博、腾讯微博等国内外最大社交网络数据,建立支持PB级数据处理能力跨地域分布式试验环境;3)音视频搜索试验床,覆盖视觉、听觉、高动态图像、全光图像等多模态信息,建立网间视音频数据的融合、互联与共享通道,实现跨网数据融合与全局化,建立具有广覆盖性的跨网一致性关联与融合的视音频检索平台。

在应用系统方面,可能的具体发展趋势包括:1)在音视频监控领域,实现全天候真实环境下的人脸、人体、车辆等典型运动目标的实时检测、跟踪与识别技术;异常行为和事件的监测、预测与评估技术。数据时空协同分析、理解与价值挖掘;2)针对社交网络领域,在PB级数据空间上,实现事件、人物、网络群体、相互关系、信息事件、情感演化等的实时搜索;3)针对医学健康领域,在隐私保护的前提下,实现电子病历、诊疗数据、专业论文、疾病症状、治疗护理推荐、医院医师等的个性化搜索;4)在物联网领域,在隐私保护的前提下,实现与人(如可搜索和跟踪指定穿戴设备的信息)、物体及状态信息(如医院剩余床位数量)、感知趋势(如温度、气压、噪音等状态信息)及行程有关的搜索。

在共性大搜索引擎的发展趋势方面,在以上具体领域成功应用的基础上,探索构建通用共性的大搜索系统,具体包括:支持涵盖了信息、人物和物体的泛在网络空间,Web 2.0和Web 3.0互联网应用模式,在大数据环境中支持对用户真实意图的理解,支持构建知识仓库,并返回智能解决方案,并支持对搜索的全生命周期的安全访问控制和隐私保护。

6 结束语

搜索引擎可以通过技术措施操控人们获得信息的范围,谁掌握了搜索引擎,谁就掌握了信息网络空间的入口,掌握了为人们提供信息甚至答案的权利,

由此产生的经济和社会驱动力日益受到各国重视, 面向网络空间的下一代搜索引擎——“大搜索”已具有迫切的需求。本文以网络空间大搜索的需求和挑战为研究目标, 提出了网络空间大搜索应具有 5S 特征, 探索了支撑 5S 特征的研究范畴, 指明了大搜索技术的发展趋势。从国际宏观上看, 目前大搜索技术仍然处于起步阶段, 应当把握切入大搜索的机会机遇, 努力与发达国家展开技术竞争, 抢占大搜索引擎这一产业的制高点, 力争掌握相关自主知识产权, 以争取在下一轮的信息革命中占据先机, 从而提高社会运转效率, 推动国家经济的健康发展。

参考文献:

- [1] HENDLER J. Web 3.0 emerging[J]. Computer, 2009, 42(1):111-113.
- [2] MANYIKA J, CHUI M, BROWN B, et al. Big Data: The Next Frontier for Innovation, Competition, and Productivity[M]. McKinsey Global Institute. 2011.
- [3] CHAU M, CHEN H. Comparison of three vertical search spiders[J]. Computer, 2003, 36(5):56-62.
- [4] HOWE A E, DREILINGER D. Savvysearch: a meta-search engine that learns which search engines to query[J]. AI Magazine, 1997, 18(2): 19-25.
- [5] PAGE L, BRIN S, MOTWANIR, et al. The PageRank Citation Ranking: Bringing Order to the We[R]. 1999.
- [6] WILKINSON K, SAYERS C, KUNO H, et al. Efficient RDF storage and retrieval in jena2[A]. International Workshop on Semantic Web and Databases[C]. 2003. 35-43.
- [7] ETZIONI O, KOK S, SODERLAND S, et al. Web-scale information extraction in knowtAll[A]. International World Wide Web Conference Proceedings[C]. 2004. 100-110.
- [8] YATES A, CAFARELLA M, BANKO M, et al. Texrunner: open information extraction on the Web[A]. Proceedings of Human Language Technologies[C]. 2007. 25-26.
- [9] WU W, LI H, WANG H, et al. Probase: a probabilistic taxonomy for text understanding[A]. Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data[C]. 2012. 481-492.
- [10] FABIAN M, GJERGJI K, GERHARD W. YAGO: a core of semantic knowledge unifying wordnet and wikipedia[A]. International Conference on World Wide Web. 2007. 697-706.
- [11] AUER S, BIZER C, KOBILAROV G, et al. DBpedia: a nucleus for a Web of open data[A]. International Semantic Web Conference and 2nd Asian Semantic Web Conference[C]. 2007.11-15.
- [12] LIU Q, WU D, LIU Y, et al. Extracting attributes and synonymous attributes from online encyclopedias[A]. International Joint Conferences on Web Intelligence and Intelligent Agent Technologies[C]. 2014. 290-296.
- [13] BROEKSTRA J, KAMPMAN A, VAN H. Sesame: an architecture for storing and querying RDF data and schema information[J]. Semantics for the World Wide Web, 2001:197-222.
- [14] ALEXAKI S, CHRISTOPHIDES V, KARVDVNARAKIS G, et al. The RDFSuite: managing voluminous RDF description bases[J]. International Workshop on the Semantic Web, 2001:1-13.
- [15] ABADI D J, MARCUS A, MADDEN S R, et al. Scalable semantic web data management using vertical partitioning[A]. International Conference on Very Large Data Bases[C]. 2007. 411-422.
- [16] WEISS C, KARRAS P, BERNSTEIN A, et al. Hexastore: sextuple indexing for semantic web data management[J]. Proceedings of the VLDB Endowment Archive, 2008, 1(1): 1008-1019.
- [17] FAN W, LI J, MA S, et al. Adding regular expressions to graph reachability and pattern queries[J]. Frontiers of Computer Science in China, 2012, 6(3): 313-338.
- [18] GUBICHEV A, BEDATHUR S, SEUFERT S, et al. Fast and accurate estimation of shortest paths in large graphs[A]. International Conference on Information and Knowledge Management[C]. 2010. 499-508.
- [19] POTAMIAS M, BONCHI F, CASTILLO C, et al. Fast shortest path distance estimation in large networks[A]. ACM Conference on Information and Knowledge Management[C]. 2009. 867-876.
- [20] WANG H, WANG H, SHAO B. Efficient subgraph matching on billion node graphs[J]. Proceedings of the Very large Data Base, 2012, 5(9): 788-799.
- [21] CAO Y, SHUAI M, WO T. Distributed graph pattern matching[A]. International Conference on World Wide Web[C]. 2012. 949-958.
- [22] LI A, JIN S, ZHANG L, et al. A sequential decision-theoretic model for medical diagnostic system[J]. Technology and Health Care, 2015, 23(s1):37-42.
- [23] LI G, OOI B C, FENG J, et al. EASE: an effective 3-in-1 keyword search method for unstructured, semi-structured and structured data[A]. Proceedings of the 2008 ACM SIGMOD international conference on Management of data[C]. 2008. 903-914.
- [24] LI X, WANG Y, ACERO A. Learning query intent from regularized click graphs[A]. Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval[C]. 2008. 339-346.
- [25] LI X, WANG Y Y, SHEN D, et al. Learning with click graph for query intent classification[J]. ACM Transactions on Information Systems, 2010, 28(3): 1-20.
- [26] MADHU G, GOVARDHAN A, RAJNIKANTH T K V. Intelligent semantic web search engines: a brief survey[J]. International journal of Web & Semantic Technology, 2011, 2(1): 34-42.
- [27] CHILTON L B, TEEVAN J. Addressing people's information needs directly in a web search result page[A]. Proceedings of the 20th International Conference on World wide Web[C]. 2011. 27-36.
- [28] HE Y, WANG K. Inferring search behaviors using partially observable markov model with duration (POMD)[A]. Proceedings of the Fourth ACM International Conference on Web Search and Data Mining-WSDM '11[C]. 2011. 415-424.
- [29] AGICHTEN E, WHITE R W, DUMAIS S T, et al. Search, interrupted: understanding and predicting search task continuation[A]. Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '12[C]. 2012.315-324.
- [30] WANG H, SONG Y, CHANG M W, et al. Modeling Action-level satisfaction for search task satisfaction prediction[A]. Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval[C]. 2014.123-132.
- [31] SADIKOV E, MADHAVAN J, WANG L, et al. Clustering query refinements by user intent[A]. Proceedings of the 19th International Conference on World Wide Web[C]. 2010. 841-850.

- [32] JETHAVA V, CALDERÓN B L, BAEZA Y R, et al. Scalable multi-dimensional user intent identification using tree structured distributions[A]. Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval[C]. 2011. 395-404.
- [33] XU J, ZHANG Q, HUANG X. Understanding the Semantic Intent of Domain-Specific Natural Language Query[A]. International Joint Conference on Natural Language Processing[C]. Nagoya, Japan, 2013. 552-560.
- [34] LIANGDA LI, DENG H, DONG A, et al. Identifying and labeling search tasks via query-based hawkes processes[A]. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[C]. 2014. 731-740.
- [35] FAN W, LI J, MA S, et al. Graph homomorphism revisited for graph matching[J]. Proceedings of the VLDB Endowment, 2010, 3(2): 1161-1172.
- [36] MAJI H K, PRABHAKARAN M, ROSULEK M. Attribute-based signatures: achieving attribute-privacy and collusion-resistance[J]. IACR Cryptology ePrint Archive, 2008:1-23.
- [37] RIVEST R L, SHAMIR A, TAUMAN Y. How to leak a secret[A]. The 7th International Conference on Theory and Application of Cryptology and Information Security[C]. Gold Coast, Australia, 2001. 552-565.
- [38] LI J, REN K, KIM K. AABE: Accountable Attribute-Based Encryption for Abuse Free Access Control[R]. Cryptology Eprint Archive, 2009.
- [39] YU S, REN K, LOU W, et al. Defending against key abuse attacks in KP-ABE enabled broadcast systems[J]. Security & Privacy in Communication Networks, 2009, 19: 311-329.
- [40] OSTROVSKY R, SAHAI A, WATERS B. Attribute-based encryption with non-monotonic access structures[A]. Proceedings of the 14th ACM Conference on Computer and Communications Security[C]. New York, USA, 2007.195-203.
- [41] PIRRETTI M, TRAYNOR P, MCDANIEL P, et al. Secure attribute-based systems[A]. Proceedings of the 13th ACM Conference on Computer and Communications Security[C]. New York, USA, 2006. 99-112.
- [42] YANG K, JIA X. Expressive, efficient, and revocable data access control for multi-authority cloud storage[J]. IEEE Transactions on Parallel and Distributed Systems, 2014, 25(7): 1735-1744.
- [43] HENGARTNER U. Hiding location information from location-based services[A]. International Conference on Mobile Data Management[C]. 2007. 268-272.
- [44] ZHAO X, LI L, XUE G. Checking in without worries: Location privacy in location based social networks[A]. IEEE Conference on Computer Communications[C]. 2013. 3003-3011.
- [45] DWORC C, NAOR M, VADHAN S. The privacy of the analyst and the power of the state[A]. IEEE 53rd Annual Symposium on Foundations of Computer Science[C]. 2012.400-409.
- [46] DE CAPITANI DI VIMERCATI S, FORESTI S, JAJODIA S, et al. On information leakage by indexes over data fragments[A]. IEEE International Conference on Data Engineering Workshops (ICDEW)[C]. 2013. 94-98.
- [47] WATERS B. Ciphertext-Policy Attribute-Based Encryption: An Expressive, Efficient, and Provably Secure Realization[M]. Public Key Cryptography-PKC 2011. Springer Berlin Heidelberg, 2011.

作者简介:



方滨兴 (1960-), 男, 江西万年人, 中国工程院院士, 北京邮电大学教授, 主要研究方向为大数据、计算机网络和信息安全。



贾焰 (1960-), 女, 四川成都人, 博士, 国防科学技术大学教授, 主要研究方向为大数据、网络信息安全和社交网络。



李爱平 (1974-), 男, 山东诸城人, 博士, 国防科学技术大学研究员, 主要研究方向为语义网络、数据挖掘和网络信息安全。



殷丽华 (1973-), 女, 辽宁朝阳人, 博士, 中国科学院信息工程研究所副研究员, 主要研究方向为物联网搜索、安全评估。