doi:10.11959/j.issn.1000-436x.2015315

物联网搜索技术综述

高云全[1,2]，李小勇[1]，方滨兴[1]

（1. 北京邮电大学 可信分布式计算与服务教育部重点实验室，北京 100876；
2. 安徽工业大学 计算机科学与技术学院，安徽 马鞍山 243032）

摘 要：随着物联网的普及和发展，物联网搜索是摆在学术界和工业界面前迫切需要解决的问题，物联网搜索因此成为当前的一个研究热点。面对越来越多的传感器以及它们所产生的数据，只有结合智能的物联网搜索，才能体现这些数据的生命力。与传统的 Baidu、Google、Bing、Yahoo 等搜索引擎不同，物联网搜索从搜索对象、物理网数据的特点（大规模的、实时变化的、高度动态的、异构的、复杂的安全环境等）到物联网搜索的架构均与传统互联网不同，这导致了物联网搜所面临的挑战将更大。由此，阐述了物联网搜索的概念、特点、相关技术，对现有的典型系统和算法进行了比较性总结，分析了目前研究中存在的问题和挑战，并展望了其未来的发展方向。

关键词：物联网；搜索引擎；综述；意图理解；情景感知

中图分类号：TP393 **文献标识码**：A

Survey on the search of Internet of Things

GAO Yun-quan[1,2], LI Xiao-yong[1], FANG Bin-xing[1]

(1. Key Laboratory of Trustworthy Distributed Computing and Service, Ministry of Education,
Beijing University of Posts and Telecommunications, Beijing 100876, China;
2.School of Computer Science and Technology, Anhui University of Technology, Ma'anshan 243032, China)

Abstract: With the popularization and development of the Internet of Things, the search of Internet of Things urgently needs to be solved in academia and industry and becomes a hot research topic. With the increasing data generated by sensors, only the search of IoT makes this data valuable. However, different from the traditional Internet search engines (such as Baidu, Google, Bing, Yahoo, etc), the search of IoT has different search objects and search architecture and data in the Internet of Things are massive, real-time, highly dynamic, heterogeneous, highly insecure. The above characteristics make search of IoT face more challenges. Firstly, the concept, characteristics, related technologies of search of IoT were summarized and presented. Secondly, several typical systems and related algorithms were described in detail. Finally, the current problems, the challenges and research prospects of this field for future were presented.

Key words: Internet of Things; search engine; survey; intention understanding; context awareness

1 引言

随着现实世界中传感器的广泛部署，互联网技术逐渐渗透到物理实体世界中，越来越多的物理实体通过传感器连接到互联网中实现信息共享，物联网在此背景下应用而生。物联网（Internet of Things）这一新的信息发展浪潮[1]引起了工业界和学术界的极大关注。物联网包含了 4 个部分：现实世界中的物理实体、用于感知物理实体状态信息的传感器、传输网络、智能处理系统。

收稿日期：2015-09-21；**修回日期**：2015-12-13

基金项目：国家自然科学基金资助项目(61370069)；霍英东基金资助项目(132032)；教育部新世纪优秀人才基金资助项目(NCET-12-0794)；安徽工业大学校青年基金资助项目(QZ201412)

Foundation Items: The National Natural Science Foundation of China (61370069); Fok Ying Tung Education Foundation (132032); Program for New Century Excellent Talents in University (NCET-12-0794)；Anhui University of Technology Youth Fund (QZ201412)

如今，物联网的应用非常广泛，已经渗入到人们的学习、工作和生活中，如物流、仓库储存、智能交通、智能家居、环境监测、公共安全等各个领域。例如人们可以坐在办公室利用手机等智能终端通过互联网对家中的家具（窗帘、冰箱、空调等）进行远程的智能控制；利用传感器对环境信息进行感知，把感知到的环境信息通过网络传送到服务器端进行分析，实现了实时、智能的环境监控和分析系统[2,3]；通过安装在道路上的红外传感器感知道路上的车流量、拥堵状况等实时的交通信息实现了智能交通系统；通过安装在包裹上的传感器物流公司可以对包裹信息进行实时监控。随着这些物联网系统的开发和应用，人们越来越需要准确、及时、智能地搜索现实世界中的物理实体信息，例如附近哪里有人少、安静的咖啡厅，到一个目的地哪条道路是最近和最畅通的，附近哪家银行排队的人最少。面向物联网的搜索服务正是在此背景下应运而生，物联网搜索[4]也成为物联网最为基础和关键的组成部分之一。

目前的搜索引擎都是针对静态信息（如网页、文档、音乐、视频）进行搜索的，如谷歌、百度等，与之不同的是，物联网搜索针对的是状态实时变化的物理实体，因此物联网搜索是实时的、动态的，两者之间的比较如表 1 所示。

表 1　互联网搜索和物联网搜索比较

维度	互联网搜索	物联网搜索
搜索对象	信息	信息、物体、人
搜索维度	一维（内容）搜索	三维（内容、时间、空间）搜索
对象状态	静态	静态和动态
搜索方案	存在性搜索	智慧搜索和挖掘
搜索要求	非实时	实时

2　物联网搜索概念

2.1　物联网搜索概念及框架

物联网搜索是指应用相关的策略和方法从物联网上获取信息（如物体、人、网页等信息），并对这些获取到的信息进行存储和有组织有序地管理，以方便用户进行搜索。物联网搜索的架构如图 1 所示，包括以下几个方面。

1) 数据采集。即对物联网空间中的数据进行采集。物联网搜索中的数据采集与获取是目的性地围绕着解答搜索要求去搜集数据的，包括语法与语义上相关的数据。不同于互联网搜索，在物联网搜索中采集到的数据类型众多，如网页、图片、音频、视频等，并且是实时的、动态变化的以及多模态的。

2) 多源数据的融合分析。物联网的搜索对象不再是单纯的网页，而是由人、机、物有机互联的复合内容。物联网搜索需要通过各种途径感知搜索者的需求，获得搜索的数据。而来源于不同物联网的

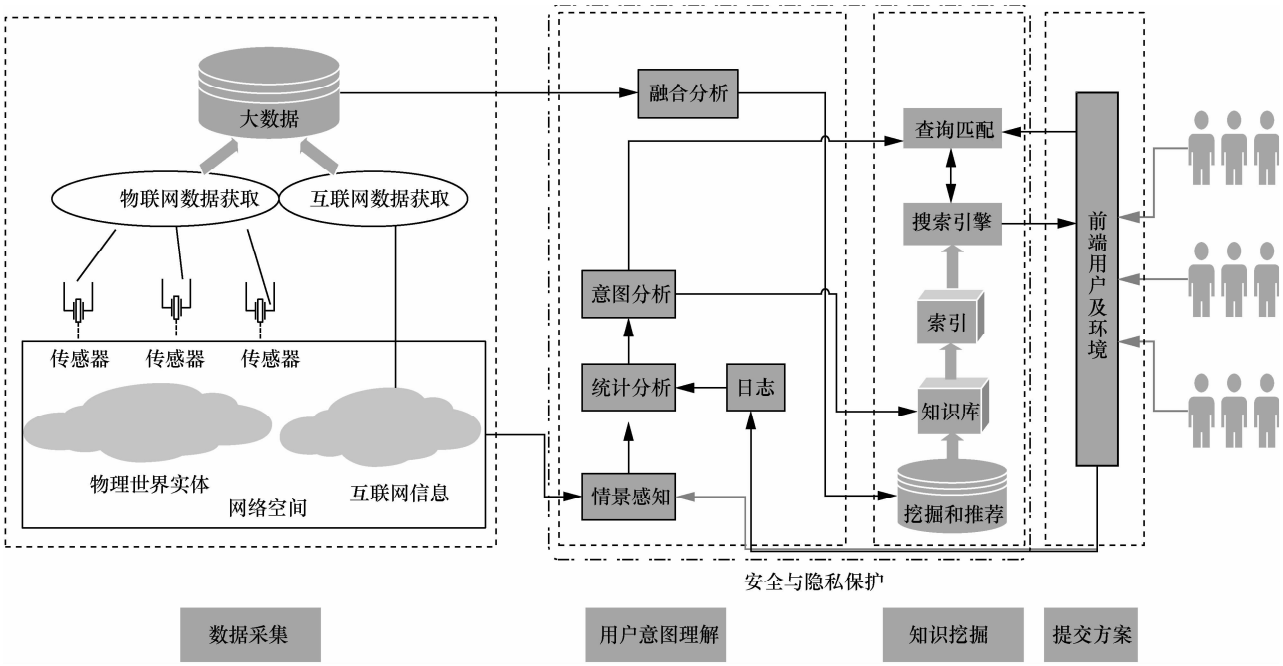

图 1　物联网搜索的系统架构

信息在性质、形式和内容上多种多样，具有多元、多属性、多维度等与传统互联网信息不同的特征，所以在物联网搜索中需要利用各种物联网终端设备实时感知用户的需求，同时对获取到的各类搜索数据进行深度分析与融合，才能准确得到所需的结果。

3) 搜索意图的理解。为了能准确地搜索到用户所需的信息，首先必须要精确理解用户的搜索意图。物联网搜索中感知用户搜索意图的渠道除了传统的文本输入之外，还可能通过物联网的各种感知设备感知用户的上下文环境，并对上下文环境信息进行分析，从而对用户的搜索意图进行更准确的理解。由于物联网具有孤岛特性，孤岛上的信息相对独立，为在物联网上实现搜索，需要将用户的搜索意图分解成若干子动作（子搜索任务），并分别在这些孤立的物联网上执行获得搜索数据。

4) 知识挖掘。基于意图理解表示和索引后的知识聚合与索引，经过快速匹配、排序等技术，形成若干个满足用户真正意图的解决方案，并通过结果评价方式给出其相关性排序。

5) 提交方案。为用户提供一个或多个智能解决方案，包括涉及用户需求的、多层面的诸多要素。通过人的参与（对用户的提问与引导、对用户需求的跟踪、对用户结果的反馈学习）来定义智慧模式，针对不同类型的问题，生成不同类型的智慧模式，用以发现符合模式的主体集合。

6) 安全与隐私保护。既要确保数据来源和推演加工结果是可信的，又要保证被搜索出的用户隐私不被曝光和恶意利用，还要能够对恶意信息进行过滤。

2.2 物联网搜索的特点

物联网搜索的对象是由传感器感知并自动生成，快速实时变化的结构化信息。而现有的互联网搜索技术如谷歌、百度等搜索引擎[5~7]，其搜索对象主要是互联网上的网页 doc 及 pdf 等由人工上传，静态或缓慢变化的非结构化内容[8]。相比互联网搜索服务，物联网搜索服务的特点如下。

1) 搜索对象的广泛性。传统搜索引擎的搜索对象是人工输入的静态内容（如网页、图片、视频等），而物联网搜索的搜索对象非常广泛，不仅包含传统互联网的搜索对象还包括动态、实时变化的实体状态信息。

2) 传感器节点的资源是受限的。由于传感器节点的电池容量、存储容量、计算能力、通信能力都是受限的，所以传感器节点必须要避免大量的、复杂的计算和频繁的通信。因此，对传感器节点进行搜索是物联网搜索面临的又一难题[9]。

3) 传感器节点是动态移动的。传感器被嵌入到物理实体中以感知物理实体的信息，因此随着物理实体的移动，传感器节点也会跟着移动，这使维护传感器的注册信息变得困难。

4) 搜索空间的广泛性。如今越来越多的物理实体嵌入一个或多个相应的传感器，用于感知物理实体的状态等信息。据估计，到 2015 年将会有数千亿嵌入 RFID 标签的物理实体[10]，时刻都会产生数以亿计的传感器数据，这将导致物联网搜索的搜索空间比传统搜索要大。

5) 数据的高度动态性。物理实体的信息随着时间和环境的变化而变化，因此传感器感知到的信息也是实时、高度变化的。例如，通过 GPS 技术测量到的旅游者位置信息可能每分每秒都在变化，因此每一个位置信息的生命周期都是短暂的。相比之下，互联网上的网页信息是静态或变化缓慢的（每隔几个星期、几个月、甚至几年才变化一次）。搜索引擎的工作过程是：通过爬虫软件每隔一定时间去爬取网页等信息的内容，然后在索引库中更新相关索引。由于物联网数据的高度动态性，所以在爬虫软件爬取相关信息并更新索引库中的索引后，在很短的时间内（几秒钟）索引所指向的信息已经发生了变化，这条最新被更新的索引很快又成为过时的。加上物联网数据的海量性，搜索引擎中会存在大量的过期索引，从而严重降低了搜索引擎的服务效率。然而如果简单地通过无限制地提高爬虫软件爬取的频率来解决索引过期的问题，这会导致通信量的急剧增大，由于通信资源的有限性，这显然不是一种可行的方法。所以，传统互联网搜索引擎的索引方法已不再适合物联网搜索，需要设计适合物联网搜索的索引方法。

6) 搜索内容的高度时空性。和传统的搜索引擎不同的是物联网搜索往往需要在某个特定空间区域中查找指定时间范围内的信息，具有很强的时空性。例如某一个特定的时间范围内不堵塞的道路、安静的餐厅或教室等信息。

7) 意图理解。结合用户请求的上下文、用户的情绪及历史偏好、被搜索对象所在环境的情景信息、时空特性等因素支持在语义上对用户搜索意图进行理解，并以统一的方式进行表示，从而明确搜

索的目标和任务。传统的搜索引擎针对不同用户的同一个搜索问题返回的结果是相同的，而物联网搜索是一种智能搜索，根据不同的用户以及所处环境的不同，返回的结果也不同。

8) 搜索语言的复杂性。传统的互联网搜索是基于关键字进行搜索和匹配的查询语言，而物联网搜索不仅需要基于关键字进行搜索和匹配还需要支持更通用的谓词来搜索物理实体的状态信息。

9) 自发的互操作。物联网系统是松散耦合的，传感器设备可以自发地相互作用。因此，物联网搜索需要一种高效的方式来处理互操作，以实现物联网搜索的规模化和实时性。

10) 用户行为的不同。互联网搜索往往关注的是分布在互联网上的信息（如网页、pdf、doc），而物联网搜索更多时候关注的是本地周围的物理实体而不是远在互联网上另一端的网页等信息。这是因为用户通常要操作和控制物理对象来实现自己的目标。例如，汤姆在办公室里查询汽车的钥匙放在什么位置，他要用哪串钥匙来开启车。

11) 智慧搜索。搜索引擎要做的工作在于如何能够给出最符合用户需求的信息。物联网搜索是一种智慧搜索，基于泛在网络获取到的数据集合，通过统一的知识与关系表示模型，在此基础上通过融合、关联、统计、推理、众包等技术进行智慧的挖掘搜索。

3 物联网搜索关键技术

虽然物联网搜索不同于传统的搜索引擎，面临着很多需要攻克的技术难关，但传统搜索引擎的一些技术仍然还是适用的[11, 12]。下面根据传统互联网搜索的一些基本技术以及结合物联网搜索本身的特点，就物联网搜索的关键技术和策略进行介绍和分析。

3.1 物联网中的数据采集和融合

由于物联网中数据的类型多以及高度动态变化性和关连性强等特点，研究快速、实时进行异构海量物联网数据获取、处理及融合的技术。

3.1.1 主动推送和被动索取（push and pull）

物联网中的实体产生数据，用户输入查询要求，输入的查询要求需要和实体产生的数据进行比较以找出符合用户需求的实体。在分布式的网络环境中，查询有3种实现方式：一种是物理实体主动将数据推送并存储在用户端，查询是在用户本地实现的，这种查询方法被称为 push（推送法）；另一种是用户输入的查询要求被发送到各物理实体，在各物理实体端进行查询，最后将符合查询要求的物理实体的数据发送给用户，这种查询方法被称为 pull（索取法）；第 3 种是采用 push 和 pull 两者相结合的方式，对于那些用户经常需要访问的数据采用 push 的方式推送到用户端，其他数据则存储在服务器端，当用户需要时再从服务器端 pull 过来。

3.1.2 发布/订阅（publish/subscribe）

对于连续查询来说，在物理实体与用户之间建立起明确的关系是非常有必要的。发布/订阅是指当用户对某物理实体感兴趣并需要经常访问时，可以向该物理实体进行订阅。当用户和特定的物理实体建立订阅关系后，针对该物理实体数据的获取过程，用户无需花费大的开销进行查询，只需要该物理实体简单地把数据定向推送到用户处即可。物理网中的数据规模和用户规模都是超大的，连续的 pull 和 push 操作会引起通信资源的巨大开销，发布/订阅技术可以减少 pull 和 push 操作，不仅节省了一定的通信资源，而且也提高了用户查询的效率。

3.1.3 数据融合技术

数据融合是指对物联网中不同类型、不同来源的数据进行关联、过滤、统计、推理、合成等获取和推演技术，发现和获取数据中蕴含的各类知识和智慧的过程。物联网中的数据具有多元化、异构性、多维度等特点，因此为了保证搜索的质量，需要在物联网搜索中对获取到的各类搜索数据进行深度分析与融合，才能得到满足用户需求的准确结果。由多个传感器融合后的信息可以更精确、更全面、更可靠，这是单个传感器无法完成的。例如，对于一个复杂的搜索任务，需要分解成若干个子任务，每个子任务在不同的子网中进行搜索。每个子网都有一个最近的搜索代理（简称邻近代理），正常情况下各个子网搜集的数据先传输到自己的邻近代理。然后，各个邻近代理把数据汇聚到一个搜索代理进行融合操作得到最终的搜索结果。数据融合包括数据层融合、特征层融合和决策层融合。

3.2 物联网中数据的存储和管理

针对泛在物联网数据的规模巨大、多维索引、实时动态更新，以及用户敏感、地理位置敏感、复杂关联分析等特性，研究支持实时动态更新、多维索引、海量动态变化数据的存储和管理技术。

3.2.1 压缩技术

压缩技术的使用减少了数据的存储量和查询的通信开销。例如，服务器通常是存储实体数据的一个压缩聚合视图。无损压缩不影响系统的基本操作。然而，有损压缩只能产生一个近似视图，因此，查询结果可能是启发式的，或者考虑用近似视图表示实体和用户的子集，然后再在该近似的视图上执行查询操作。例如，对于有损压缩，一个极端的例子是仅仅传送或存储一比特的信息，这意味着实体产生的数据可能已经发生了变化，因此 pull 操作只需考虑发生改变的实体。

3.2.2 中间件（mediators）

中间件是一个概念性元素，逻辑上位于实体和用户之间，是目前使用较多的一种数据集成方法。中间件通常维持了实体的聚合视图，发送到中间件的查询请求无需向所有实体索取数据即可完成查询。中间件可以是集中式的也可以是分布式的（例如在一个分层的中间件中，上级中间件 supermediator 维持着一个下级中间件 submediators 的聚合视图）。

3.2.3 云计算

云计算（cloud computing）是基于互联网的相关服务的增加、使用和交付模式，通常指通过网络以按需、动态易扩展的方式获得服务和资源。云计算的基本理念是，将大量计算和存储资源通过网络连接起来并进行统一管理从而形成一个云。当用户需要服务时，通过网络向云发出请求，云提供相应服务并将结果返回给用户。云计算是分布式处理(distributed computing)、并行处理(parallel computing)和网格计算(grid computing)的发展。物联网具有数据海量性、高度动态实时变化性、异构性等特点，这必将导致在物联网搜索中对计算资源的需求提高，云计算为此提供了技术支持。

3.2.4 语义技术

通过传感器感知到的信息是冗余的和不确定的[13]。信息的不确定性包括异构性、不连续性、不准确性、不完整性、不一致性等。其中，异构性尤为突出，通常表现为感知数据在性质、类型、内容和表达形式等的不同。异构性给信息处理、整合和描述增加了难度。物联网的异构性（如感知设备、数据格式、标准等）导致了感知信息也是异构的。因此，如何实现传感器设备的自动部署、发现和接入是物联网搜索面临的关键技术之一。此外，随着传感器设备数量的快速增长，物联网中的数据规模也在飞速增长，导致了大量的冗余信息的产生。这给数据的传输、存储、处理带来了很大的挑战。另外，在多数情况下，人们并不需要获取全部的感知数据，而只需获取语义信息或事件。如何从物联网的海量数据中获取语义信息和知识是物联网搜索的一项关键技术。

针对以上问题，语义技术提供了一种可行的解决方法。语义技术是解决不确定性和冗余性的关键技术。

3.3 物联网搜索中用户意图的理解与表示

3.3.1 情景感知意图理解

情景也即上下文，是指反映实体所处环境特征的信息，例如实体所处的空间、时间、温度等。情景感知[14, 15]是指收集情景信息，并对情景信息进行智能处理的过程。准确理解用户的搜索意图是提高搜索质量的前提和基础。物联网搜索中感知用户搜索意图的渠道除了传统的文本输入之外，还可能通过物联网的各种感知设备感知用户的上下文环境，依据情景信息，更准确地理解用户的搜索意图。例如，通过将位置信息、查询信息、社交关系等因素引入到背景知识构建算法中，并利用该相关性以及各种时空情景实时判断用户的查询语义。

用户意图理解有：基于时空的用户意图理解、基于形体动作的用户意图理解、基于情感分析的用户意图理解、基于统计分析的用户意图理解，交互式用户意图理解。基于时空的用户意图理解是指用户在查询时并没有给出时间和空间，但查询过程会自动去感知查询的时间和空间意图。基于形体动作的用户意图理解是基于用户的形体的一系列动作，如手势、表情、肢体语言等来推测用户的搜索意图。基于情感分析的用户意图理解是通过分析用户的情绪来选择与其个人风格、偏好相符的结果信息。基于统计分析的用户意图理解是指根据用户的历史关键词记录，选取搜索结果记录等历史偏好的统计信息，来理解用户的搜索意图。交互式用户的意图理解是指通过与用户的人机交互行为来理解用户的意图。

典型的情景感知的框架包括：情景信息的采集、建模和处理。情景信息采集即通过传感器或人机交互方式获取物理实体的情景信息。物联网有着海量的传感器，不同的传感器采集到的信息以及信息的表示方式都不同，情景信息建模是对这些数据

进行标准、统一的描述。情景信息建模有2个层面的，一个层面是形式上的统一，另一个层面是语义上的统一。形式上的统一是指对不同情景信息所采用的描述方式进行统一化和标准化，如关键值模型，语义上的统一，本体论模型就属于此。本体论模型通过本体论的知识表示解决情景信息的语义理解和互操作问题。情景信息处理是指通过可利用的情景信息推出新的知识，以便对实体有更好、更深的理解，它是一个从已知的情景信息的集合推导出一个高级情景信息的过程。情景信息推理分为3步。第1步是情景信息的预处理，即传感器的原始数据进行数据清洗。传感器硬件的性能较低以及通信资源受限，这导致从传感器收集到的数据精度不高，存在噪声甚至出现数据丢失。因此，数据需要通过填充缺失的值，去除异常值，验证情景信息的方式进行数据清洗。这一任务已广泛应用到了数据库领域、数据挖掘领域和物联网领域中。第2步是传感器数据的融合。这是一个融合多个传感器数据以产生一个更精确、更完全、更可靠信息的过程。在物联网中由于传感器的规模巨大并且产生的数据种类众多，因此数据融合对物联网来说显得非常重要。第3步是情景信息的推理（也即情景决策）。通过对低层次的情景信息进行推理可以得到高层次的情景信息。有许多不同的情景推理的决策模型，如决策树、朴素贝叶斯、隐马尔可夫模型、支持向量机、*k*-近邻、人工神经网络 D-S、基于本体论的、基于规则的以及模糊推理等。情景信息推理有以下6类[16]：监督学习、无监督学习、规则、模糊逻辑、本体推理和概率推理。

1) 监督学习：这类技术要求首先收集训练样本，接着根据所期望的结果对样本进行标记，然后推导出一个函数，该函数通过使用训练数据产期望的结果。这一技术在移动电话[17]的感知和行为识别[18]中得到了应用。决策树、人工神经网络、贝叶斯网络、支持向量机都属于监督学习型。

2) 无监督的学习：这类技术能够在未标记的数据中发现隐藏的结构。由于没有训练数据，所以没有错误或奖励信号来评估一个潜在的解决方案。聚类技术如 *K*-最近邻被广泛应用在情景感知推理中。

3) 规则推理：这是最简单、最直接的方法。规则通常用 IF-THEN-ELSE 这种格式来表示。对低级的情景信息进行规则推理可以产生高级别的情景信息。近年来，规则已被大量应用于本体推理中[19~21]。例如，MiRE[22]是一个针对情景感知移动设备的小规则引擎。大多数的用户偏好使用规则进行编码。PRIAMOS[23]使用语义规则对情景信息实施注释。

4) 模糊逻辑：与精确推理不同，模糊逻辑是一种近似推理。模糊逻辑类似于概率推理，但是它的值表示的是相似度而不是概率。在传统逻辑理论中，真值是0或1（即假或真），而在模糊逻辑中，真值不再是非真即假，可以是部分的真。由于很多真实世界的因素并不是绝对的，因此模糊逻辑更自然地表达了真实世界。模糊推理通常不能作为一个独立的推理技术，而是用来补充其他技术如基于规则的推理，概率或本体论。文献[24,25]使用了模糊逻辑表达情景信息。

5) 基于本体的推理：本体推理是基于描述逻辑的，它是形式化的知识表示逻辑系列。常见的本体推理的描述语言是 RDF(S)[26]和 OWL(2)[27]。本体推理的优势在于它可以和本体建模进行很好地集成。缺点是本体推理无法找到丢失的值或者模糊的信息，这是统计推理擅长的。本体推理已经被广泛应用于各种领域，如活动识别和混合推理[28, 29]和事件检测。

6) 概率推理：概率推理即根据问题有关的实事的概率做出决定。它可以用来结合2个不同来源的传感器数据。此外，对于情景感知中产生的冲突问题也可以用概率推理加以解决。Dempster-Shafer 是基于概率逻辑的，允许不同的证据相结合来计算一个事件的概率，是常用的活动识别传感器数据融合。

3.3.2 建模（models）技术

通过建模技术可以无需进行交流而直接推断出相关信息。用户可以通过模型实现只索取其感兴趣的物理实体数据。实体通过模型实现只推送给对其感兴趣的用户处，模型是基于历史信息创建的。建模技术可引发启发式查询解析或者用来确定用户和实体集，确保随后的推送和索取操作能够获得精确的结果。

3.4 面向物联网搜索的知识挖掘

3.4.1 倒排索引（inverted index）

倒排索引是一种数据结构，其显著提高了搜索的效率。倒排索引解决了如何根据属性值高效地查询记录。倒排索引表中存储这属性值以及各记录的地址[30]。由于是通过属性值来确定记录的位置，因此称其为倒排索引。

3.4.2 评分与排名（scoring and ranking）

评分是关于实体与查询相关度的一个标量值比例，排名则是依据评分得到的一个实体排序。对实体进行评分和排名首先可以为用户提供最匹配的实体，其次，通过对排在前面，匹配度高的实体先进行推送和索取提高了搜索效率。然而，如何统一化、标准化评分是其实现的关键前提。

3.4.3 Top-*k* 查询（Top_*k* query）

在实际的操作中，如果将所有的匹配结果都返回给用户，这不仅浪费了计算和通信资源，对用户来说也是没必要的。因此，用户进行查询时往往无需返回所有匹配的实体，而只需要返回最相关的 *k* 个结果，即 Top-*k* 查询[31]。对于 Top-*k* 查询，有时可以直接查找到最相关的 *k* 个结果，而无需考虑所有的实体，这比首先找到所有可能的匹配结果，然后再返回最相关的 *k* 个结果的这种方法大大提高了查询效率。

3.4.4 预测技术

预测是对客观事物的发展规律和趋势进行的预计与推断。预测的目的是揭示事物发展规律，预测事物未来的发展趋势，并使人们可以利用事物的规律和趋势对事物进行控制，为人类提供服务。例如，文献[11]利用卷积和傅里叶变换计算有关人类活动的周期性规律，从而达到对传感器未来的状态进行准确的预测，这不仅可以节省通信资源还可以提高物联网搜索的高效性和准确性。

预测有 2 类：定性预测和定量预测。定性预测是指凭借直观，依靠经验，通过分析对事物的未来进行的一种预测。定量分析是指通过数学工具进行统计分析的方法对事物进行的一种预测。定量预测具体方法有：回归分析法和时间序列分析法。回归分析法是一种根据事物发展的因果关系进行的一种预测。回归分析主要研究引起事物变化的各因素之间的相互作用以及各因素与未来状态之间的统计关系。具体可以通过机器学习法建立预测模型，典型的机器学习方法包括：决策树方法、人工神经网络、支持向量机、正则化方法、近邻法、朴素贝叶斯（属于统计学习方法）等。时间序列分析法也叫趋向外推法，它是根据历史数据，对事物的发展规律进行分析推理。时间序列分析法把发生的时间按照时间进行排列，然后通过趋势外推进行预测。时序分析研究的是预测目标和时间之间的演化关系，因此时间序列分析法是一种定时的预测技术。物联网搜索的搜索空间是大规模的，数据也是高度动态的，而预测技术的使用不仅可以节省通信资源，还可以提高搜索的效率以及准确性。

3.4.5 协同搜索技术

当今社会分工越来越细协作越来越紧密，分工与协作在人类社会发展的历程中越来越显得重要。随着物联网技术的发展，搜索任务也越来越复杂，因此搜索同样也需要分工与协作，于是协同搜索应用而生。协同搜索是指通过众多参与者的有序分工与协作共同完成一个搜索任务。物联网的搜索数据具有异构性、多元性、多模态性、多属性和多维度性等特征，使物联网搜索比传统的互联网搜索面临的问题更多，形式更复杂，任务更艰巨，协同搜索可以降低物联网搜索的复杂性并且提高搜索的高效性和准确性。

Web 搜索通常是用户单独进行信息搜索的行为，不同的用户输入相同的搜索词，将得到相同的搜索结果。因此，人们提出了协同搜索技术，以支持多用户高效协作搜索。例如，在医学以及军事指挥等特定领域的信息搜索中，搜索任务可以通过分工协作的方式协同完成，搜索结果可在成员之间进行共享。开始使用协同搜索的是医学视频检索等特定研究领域。Smyth 等[32]在 2003 年在第 18 届国际人工智能联合大会（IJCAI03）上发表了论文“Collaborative Web Search”中，第一次提出了协同搜索的概念。该文基于元搜索引擎设计和开发了一个协同搜索的原型系统 I-SPY。随后，Smyth 等研究小组[33,34]又继续发表了多篇协同搜索方面的论文。Morris 研究小组[35]开发了 TeamSearch、S3、Cosearch、SearchTogether 等协同搜索系统。

3.5 面向物联网搜索的安全与隐私保护

面向物联网搜索的安全与隐私保护包含以下几个方面：隐私保护、访问策略的隐藏、安全属性匹配、数据融合的安全性问题。

物联网中产生的数据需要进行隐私保护处理，隐私保护技术有以下 3 类[36]。

1) 基于数据失真（distorting）的技术：采用添加噪声、交换等方式对原始数据进行干扰处理，但同时保持某些关键数据属性的不变性。

2) 基于数据加密的技术：是指在数据挖掘过程中采用加密技术对敏感数据进行隐藏。该方法多用

于分布式环境中，如安全多方计算 SMC[37, 38]（secure multiparty computation），即站点之间通过协议完成计算后，各自都只知道自己输入的数据和通过加密技术对其进行计算后的最终结果。

3) 基于限制发布的技术：根据情况有选择地对原始数据进行发布、例如不发布精度较高敏感数据或者发布精度较低的敏感数据、数据泛化（generalization）[39, 40]等实现隐私保护。当前基于限制发布技术的研究主要集中在“数据匿名化”。包括 *L*-diversity、*k*-anonymity、*T*-closeness。

在搜索过程中用户对数据的存取策略需要对执行搜索的第三方进行隐藏，然而在大规模的物联网搜索模式下，执行搜索的第三方可以通过监控搜索平台来统计授权用户的历史查询记录，推测用户的个人查询模式及全体用户的全局数据存取策略，进而获知热点数据的分布情况及授权用户的权限等级等隐私信息。从物联网搜索角度来看，访问模式隐藏是为了防止搜索平台对用户访问模式的挖掘。PIR（privacy information retrieval）协议[41,42]将用户的查询请求通过一个矩阵变换构造出 $N-1$ 个与其不可区分的伪查询，使攻击者对用户的真实意图无法准确把握，从而实现在数据搜索平台上用户访问策略的匿名。

在物联网中，不能让用户随意搜索未授权的数据，进行搜索前需要对授权用户和搜索内容的相关信息进行验证即安全属性匹配。根据使用的数学理论安全属性匹配主要分为基于交换加密的匹配协议、基于伪随机函数的匹配协议、基于线性多项式的匹配协议以及授权和基于策略的匹配协议。

1) 基于交换加密的匹配协议。Agrwal 等[43]提出了一种建立在交换加密基础上的 PSI(private set intersection) 协议，交换加密函数具有性质：$E_{k1}(E_{k2}(P))=E_{k2}(E_{k1}(P))$。该协议建立在 DDH (decisional Diffie-Hellman)之上，协议的复杂度是线性的。此外，该协议是一种单向的交集计算协议，无法抵御恶意攻击。

2) 基于伪随机函数的匹配协议。为了使协议能够抵御各种攻击并且具有更高的运行效率，Jarecki 等[44]提出使用一个承诺密钥并用伪随机函数进行加密的 PSI 协议。该协议规定伪随机函数的输入域必须是多项式的。

3) 基于线性多项式的匹配协议。Freedman 等[45]提出的 FNP 协议是基于多项式估值和加法同态加密的。在该协议中，通过将数据集中的数据作为多项式的根构建一个多项式，然后对多项式中的系数进行同态加密，该协议的复杂度是线性的。但是，在该协议中只有客户才能知道交集，服务器是无法获得任何信息。该协议无法防止恶意攻击，适合于半诚实模型。为了解决这一问题，提出了 2 个改进的协议：一个是适用于一方半诚实而另一方恶意的场合；另一个则是适用于双方都恶意的场合。

4) 授权和基于策略的匹配协议。在普通的 PSI 协议中，攻击者通过将一些虚假的、猜测的元素插入集合中并和对方进行匹配的方式得知该元素是否存在于对方。为了解决这一问题，提出了基于授权的 PSI 协议，通过可信第三方授权，保证双方所交互的元素都是可信的。另外一个问题是当交集大小为客户端的输入大小，这种情况下对客户端而言是不安全的，为了解决该问题，先得到交集的大小，随后根据客户端的要求和条件决定是否继续下一步的匹配。在文献[46]中，Stefanov 等提出了一个加强的基于策略的 PSI 协议，该协议的属性是经过授权的，此外该协议证明了在恶意模型中的安全性。

物联网搜索需从多种类型的网络中读取数据，并服务于用户，然而这种跨网模式的搜索及数据融合必须以数据安全性为基础。安全数据融合的目的是为了保证最后得到的融合结果是正确且可接受的[47]。当前的安全数据融合有以下几类。

1) 同态加密机制安全数据融合方案。同态加密机制源于私密同态[48]，是建立在代数运算基础上的。同态加密是直接在密文上进行操作的一种机制，是端到端的一种加密方式，中间节点不需要加解密，可以实现求和、乘积的融合操作，保证了数据机密性。同态加密由于是直接在密文上进行操作，减少了计算代价并且延长了网络的生存时间，保证了数据的端到端安全。同态加密算法的例子有：Ferrer[49]等提出的一种新的私密同态算法；Mykletun 等提出的支持简单求和运算的 AHE[50]算法和基于椭圆曲线的加法同态私密算法 ECEG 算法[51]；Girao 等[52]提出的 CDA 算法；Mlaih 等[53]提出的一种复合运算的算法；Rodhe 等[54]提出的一种 *n* 层安全数据融合算法（*n*-LAD）；Bahi 等[55]提出了端到端的基于椭圆曲线加密的安全数据融合算法；

Zhang 等[56]提出的算法（b-pha）。

2) 隐藏真实数据的安全数据融合方案。基于隐藏真实数据安全数据融合的算法的例子有：Cam 等[57]提出的基于模式码的能量有效的安全数据融合算法（ESPDA）以及基于参考数据的安全融合算法（SRDA）[58]；He 等[59]提出隐私保护算法（PDA），该算法采用了数据切分重组和扰乱技术来保护数据的机密性；Zhang 等[60]提出的 GP2S 算法；Li[61]在 SMART 方案的基础上进行改进而得到的 CACR 算法；Groat 等[62]提出的 KIPDA 算法；Li[63]和杨庚[64]分别提出了 EEHA 算法和 ESPART 算法；Bista 等[65~67]提出的一些新型的算法。

3) 监督和信誉机制的安全数据融合方案。基于监督和信誉机制安全数据融合的算法的例子有：Du 等[68]提出的算法 WDA；Gao 等[69]在 WDA 协议的基础上进行扩展提出了算法；Ozdemir 等[70,71]提出 SELDA 算法以及对 SELDA 算法进行改进而提出的 RDAT 算法；Vu 等[72]提出的算法 THIS；Bohli 等[73]提出的一种安全数据融合算法。

4) 数字签名安全数据融合方案。数字签名安全数据融合的算法有：Mahimkar 等[74]提出的一种适用于分簇型 WSN 完整性数据融合算法 SecureDAV；Yang 等[75]提出的算法 SDAP；Li 等[76]提出的一种高效可靠的基于身份认证的安全数据融合算法。

4　典型的物联网搜索系统

4.1　Snoogle/Microsearch

由美国弗吉尼亚州立大学（彼得斯堡）的研究人员设计的 Snoogle 系统中[77, 78]，对实体的描述是以一组关键字（文本信息）的形式存储在传感器节点中。Snoogle 的系统架构如图 2 所示。

该系统的思想是用关键字对连接到物理实体的传感器进行描述，用户通过关键字查询匹配的目标物理实体，系统将返回查询相匹配集中最相关的 K 个实体。该系统由 2 层 mediator 组成。底层的 mediator 称为索引点（IP, index point），每个 IP 维护管理一个特定范围的传感器（也即一个传感器属于唯一的 IP）。顶层的 mediator 称为关键索引点（KeyIP, key index point），关键索引点维持着整个网络的聚合视图。传感器传送变化的文本描述信息到 IP 节点，所有的 IP 节点在把信息传送到关键 KeyIP 节点。当用户查询的是某个特定 IP 中的实体信息时，用户可以直接向该 IP 节点发送查询请求。用户也可以向一个 KeyIP 节点发送查询，查询全局范围内的实体信息。为了实现高效查询，IP 和 KeyIP 都使用了倒排索引技术。大部分的研究工作致力于如何在典型的传感器节点的缓慢、页面结构的闪存中维持索引。移动节点的搜索是通过在 IP 节点之间使用交换协议实现的。IP 节点周期性地发送信标信息，检测传感器的存在。当检测到传感器出现或者消失时，IP 节点将更新其索引项，并通知 KeyIP 节点。为了压缩通信，提高通信效率，采用了 Bloom filter 表示一个关键字集合。Bloom filter 是一个 m bit 的二进制的向量。通过相互独立的 n 个散列函数将一个元素映射到 Bloom filter 向量中的 n bit，映射位置被设置为”1”。为检查一个关键字是否属于 Bloom filter 集合，需要对这个关键字应用 n 次散列函数，如果所有映射位置都是”1”，则认为是属于

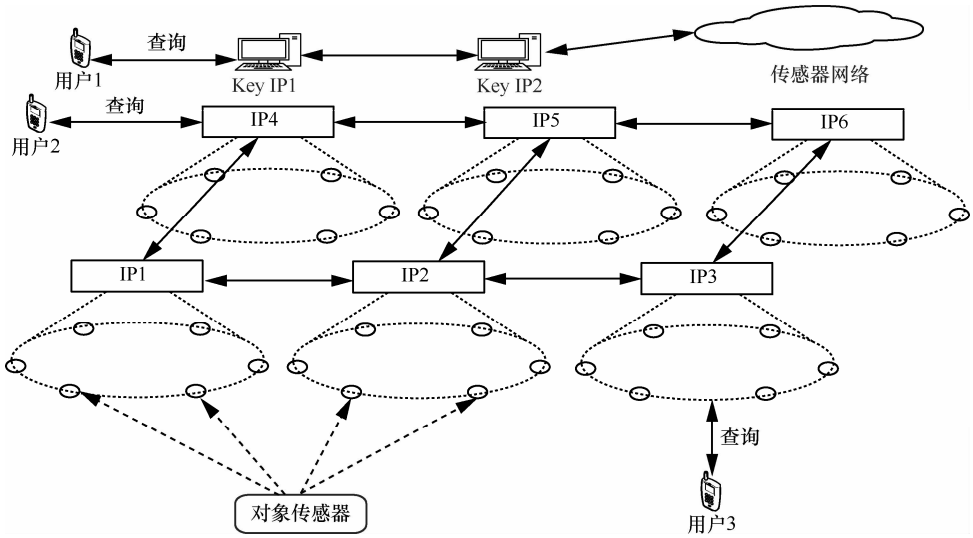

图 2　Snoogle 系统架构

Bloom filter集合。在判断一个关键字是否属于一个Bloom filter集合时有可能会把本不属于该Bloom filter集合的元素误认为属于这个集合，但本身是属于集合中的元素是不会出现漏判的。当处理查询时，传感器将按照其包含的关键字的数量进行排序。为了在各IP之间标准化排名，关键字在一个IP包含的所有传感器中出现的总频率被纳入了排名计算中。进行局部查询时，查询请求直接发送到一个本地IP，IP利用倒排索引技术对匹配集合中传感器进行排名并返回排名最前的k个结果。对于全局查询而言，查询请求将被传送到KeyIP。KeyIP在计算全局最匹配的k个结果时并没有索取所有IP节点的全部匹配结果。首先，发送查询请求给所有IP并返回每个IP最高排名的传感器节点，并对这些节点进行排序，得到一个有序列表并存储在KeyIP中的全局列表中。在全局列表中排名最高的传感器作为全局查询结果的最高排名传感器返回给用户。为了得到排名第二的传感器，KeyIP继续发送查询请求给所有IP节点，但仅返回比全局排序表中排名第二的传感器排名分数高的传感器，并把这些结果依序插入到全局排名表中，这时全局排序表中排名第二的传感器就是所要找的结果再返回给用户。继续以此类推，直到向用户返回了排名最高的k个传感器。

该系统的局限性有2点：首先，尽管采用了push方式及时地推送传感器数据到IP和KeyIP来解决元数据动态变化的问题，但这一方式显然无法应用于大规模的网络环境，因此该系统不支持动态的数据搜索，仅支持静态数据搜索和伪静态数据搜索；其次，由于KeyIP集中管理整个网络的完整视图，对于每一个全局查询KeyIP都需要查询索取所有IP节点，因此该系统不适合全球化的搜索。而Top k算法在减少通信开销的同时也产生了大量的消息；最后，Bloom filter压缩算法的使用导致查询结果是不精确的。

4.2 MAX

由新加坡国立大学的研究人员设计的MAX[79]，其系统架构如图3所示。在MAX系统中用标签代替传感器对物理实体进行感知，与Snoogle相似的是标签中存储了对物理实体的文本描述信息。用户通过输入一组关键字进行查询，MAX返回匹配度最大的前k项给用户。MAX的一个目标是使用户很容易找到目标实体。为此，MAX采用了3层结构的mediators（中间件）组织形式。最底层子站代表一个可移动的目标（如一张桌子、一个书架），在子站上可以布置移动标签实体。中间层基站代表一个区域（如一个房间），负责管理一定范围内的所有子站；最上层的MAX服务器管理所有基站。当知道目标实体属于哪个基站和子站后就很容易定位该目标实体。在一个原型系统中，RFID标签被嵌入到了目标实体中，可以进行短距离的通信，子站和基站是传感器节点，MAX服务器是一台工作站主机。

该系统的查询方式是采用pull方式而不是push方式。MAX服务器中维持着一个基站和其位置信息的目录，因此用户可以选择在哪个基站或位置进行查询。查询请求的一组关键字被发送到所选择的基站，基站再向其范围内的所有子站广播查询信息，子站再向其范围内的所有标签广播查询信息。标签在收到查询请求后进行查询匹配，然后把匹配结果返回给子站，子站再把获取到的查询结果返回给基站，基站对匹配结果进行排序。最后，基站将返回匹配度最高的k个标签给MAX服务器和用户。

MAX采用了pull方式获进行查询，所以无需维护和更新索引，这适合于移动及内容经常改变的查询。尽管该系统是针对伪静态的元数据设计的，但该系统可以扩展成基于内容的搜索。然而，该系统的缺陷是消息需要广播到每一个子站和标签，这导致了通信开销大，使其不适用于大规模的网络环境。

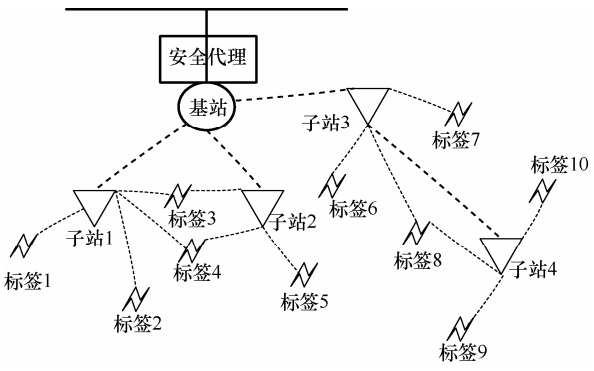

图3 MAX系统架构

4.3 OCH (objects calling home)

由瑞士苏黎世联邦理工大学和德国都科摩通信实验室设计的OCH[80]是一个提供寻找失物的系统，是一种实体目标定位系统。不同于存储实体描述信息的系统，OCH系统中的每个实体都贴上了一个电子标签，该电子标签含有实体的身份信

息。使用移动传感器检测物理实体的存在性和身份信息。在一个原型系统中，装有蓝牙技术的移动电话作为目标传感器，目标实体装有小的蓝牙电子标签模块。用户通过身份标识查询一个目标实体，系统将返回丢失物体的近似位置。在查询过程中，用户可以指定一个超时时间 t 和预算 q。移动目标传感器将对丢失的物体进行持续查询一直到超时时间 t，预算 q 限制了查询时发送的消息总数。该系统是一个基于身份信息查询实体位置的系统，这似乎与基于内容查询的系统不同。然而，假设有这样的一个传感器，它输出的内容是最后进入该传感器范围标签的身份标识信息，因此可以把该系统看成是一种特殊的基于内容的搜索系统。

OCH 系统架构如图 4 所示。在 OCH 系统架构中，移动电话连接前端的用户感知功能和后端的基础设施。移动电话的感知功能包括感知实体的存在、移动电话的位置以及有关丢失物体的其他情景信息。此外，系统架构中还包括了具体的应用程序服务：如关联性注册、用户位置分析、用户数据库。关联注册有 3 个主要目的：首先，跟踪物体和他们拥有者之间的关联；其次，物体传感器的用户允许其他用户维持一个跟他们有关的物体传感器的集合（例如，在家或者办公室安装的物体传感器；Bob 的移动设备和 Alice 建立了关联）；最后，用户和用户的关联使某些特定对象具有组访问权限。用户位置分析是根据过去所在的位置信息进行统计分析，这使用户搜索物体时可以优先考虑某个位置范围，查询无需发送到所有的目标传感器，具体策略如下。1）物体可能在其最后被看到的位置附近；2）物体可能在其所有者最近常访问的位置附近；3）物体可能在其所有者最常访问的位置处。当用户执行查询时，查询请求信息包含目标实体，上述优先策略、超时时间 t、预算 q。执行查询时将根据优先策略创建一个实体可能出现的位置排序列表，在这一过程中不需要跟任何传感器进行通信，节省手机电池等资源提高了系统的可扩展性。随后根据排序列表依次查询相应传感器直到目标实体被找到、预算 q 耗尽或者查询时间超过时间阈值 t。

用户数据库中存储了应用数据，如之前的某些对象的报告以及对象离开一个传感器范围时的相应处境信息等。查询服务整合了以上所有的部分。查询服务包括本地查询、全球查询和指定范围查询。本地查询指的是在单部手机覆盖的范围内进行的查询；全球查询是指使用移动蜂窝网在全球范围内进行查询；指定范围查询通过历史数据预测出某个可能的范围，然后在该范围内进行查询。

该系统的主要优点是：由于优先策略的使用，查询时只需要和少数目标传感器进行通信，建立链

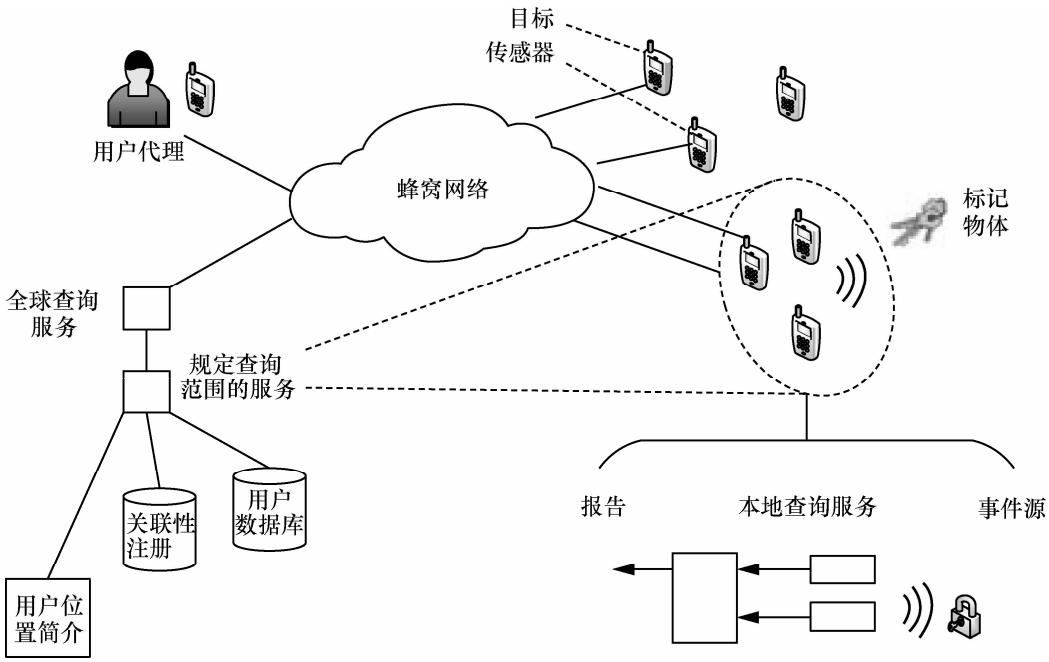

图 4　OCH 系统架构

接，使该系统能够适应大规模的网络环境。系统的缺点是：除非 q 设置为无限大，否则可能出现目标物体确实存在但系统却搜索不到的情况；该系统是一个基于身份信息查询位置的搜索系统，不能扩展成基于内容的搜索系统；由于需要计算优先策略，因此，模型的计算开销很大。

4.4　GSN (global sensor network)

由瑞士的洛桑联邦理工学院和爱尔兰国立大学设计的 GSN[81]系统将异构的传感器和传感器网络通过 Internet 进行互连，支持在全球传感器数据流集合中进行同质数据流的查询。GSN 的系统架构如图 5 所示。虚拟传感器是 GSN 提出的一个重要的抽象概念。一个虚拟传感器可以表示一个物理传感器或者一个虚拟实体。一个虚拟传感器可以有一个或多个虚拟传感器的数据流作为输入，经过处理后产生一个输出数据流。GSN 支持发现虚拟传感器以及把一个或多个虚拟传感器互联起来形成一个新的虚拟传感器。GSN 中的每一个虚拟传感器都有唯一的身份标识，并使用元数据对虚拟传感器进行描述。因此，用户可以通过标识符、关键字、位置等静态元信息进行查询。GSN 系统的局限性在于：首先，不支持基于内容的查询；其次，虽然可以使用数据流查找给定输出值的传感器，但这必须要和所发现的传感器进行通信，因此 GSN 不适于大规模的传感器网络系统中。

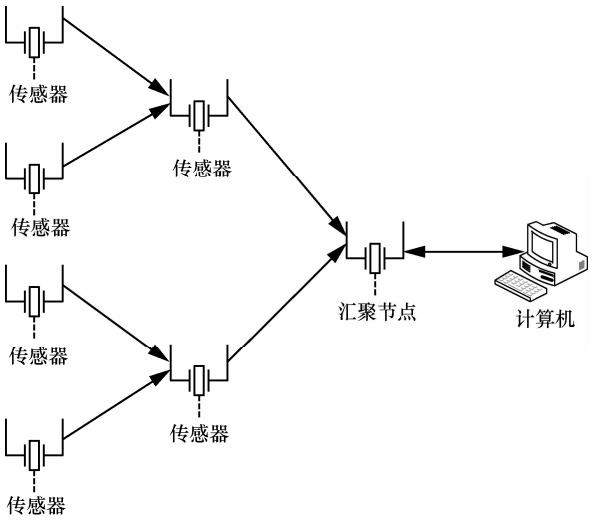

图 5　GSN 系统架构

4.5　Dyser

4.5.1　Dyser 概述

由瑞士苏黎世联邦理工大学、德国吕贝克大学以及德国都科摩通信实验室设计的 Dyser 是一个物联网实时搜索引擎。该系统不仅能够查询物理实体的静态信息，还能根据用户指定的当前状态实时地搜索物理实体。Dyser 的系统架构如图 6 所示。当前的通用搜索引擎无法搜索传感器产生的实时动态变化的数据流。为此，首先 Dyser 将物理实体以及传感器抽象为 Web 页面，以便能使用通用搜索引擎对其进行索引。Web 页面包含了传感器类型等静

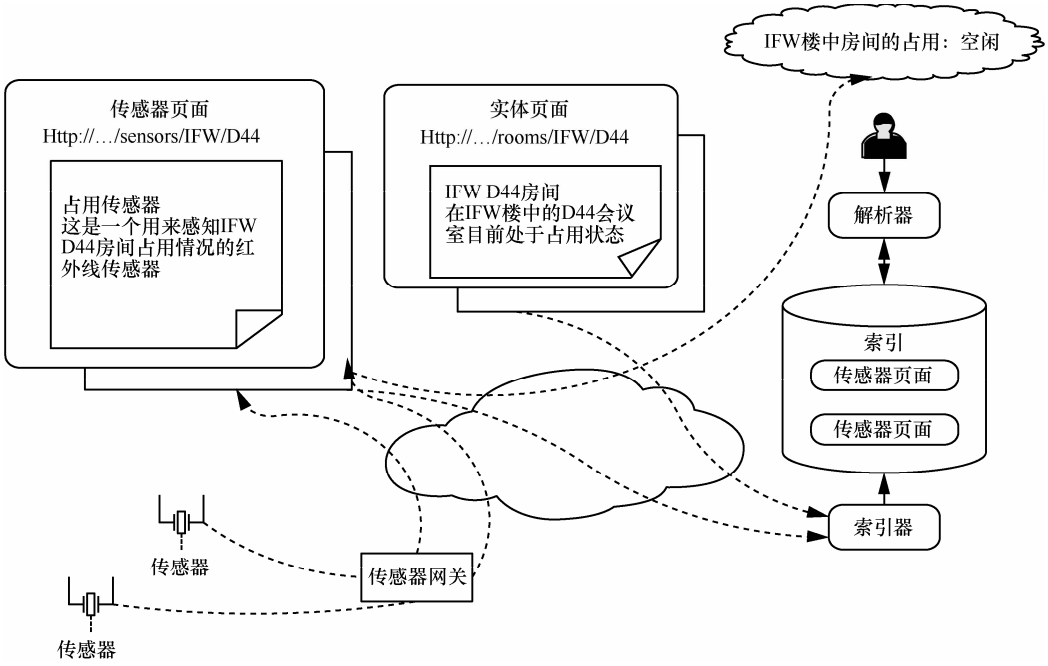

图 6　Dyser 系统架构

态的文本描述信息以及动态变化的状态等元数据信息。传感器页面和实体页面是一种多对多的关系，并通过超链接进行关联。Dyser 使用通用搜索引擎为传感器页面和实体页面建立索引库。其次，Dyser 利用过去的实体状态数据建立预测模型对实体当前和以后的状态进行预测，从而实现对实时动态的状态信息进行搜索。Dyser 提出了 3 种预测模型；聚集预测模型（APM）、单周期预测模型（SPPM）和多周期预测模型（MPPM），根据不同场合可以选择不同的预测模型。由于传感器数据的高度动态变化的特性，这可能会导致刚建立完索引后，传感器数据的内容又发生了变化，这时此索引就是一个过期的索引，不代表传感器当前的内容，但是通过预测模型的引入可以推测出当前传感器的内容。因此，预测模型的引入提高了搜索的效率。预测模型包含在虚拟传感器或虚拟实体的 Web 页面中。搜索请求是由静态请求信息和动态请求信息 2 部分构成的。搜索时首先按照静态请求查找出匹配的实体页面，然后利用预测模型计算出实体页面与搜索请求中动态属性匹配的概率，并按照匹配概率由高到低对实体页面进行排序，最后对有序列表中实体的状态信息和搜索请求中的动态请求信息进行匹配得出最终的查询结果。当一个实体所有的状态都符合要求时，就输出该实体，重复操作，直到返回足够的匹配实体。

4.5.2　Dyser 中的预测模型

传感器用式（1）表达

$$S:\tau \mapsto \upsilon \tag{1}$$

其中，τ 代表离散的时间，υ 代表有限的、离散的传感器状态的输出值集合。例如一个监控房间占用情况的传感器，$\upsilon=\{free,occupied\}$。预测模型表示为

$$P_{s,t^0,t^1}:\tau\times\upsilon\mapsto[0,1] \tag{2}$$

在式（2）表示的模型中，t^1 表示模型创建的时间，t^0 表示最早的那个传感器输出值的时间。$TW=t^1-t^0$ 代表一个时间窗口，模型是依据过去 TW 时间内传感器 S 的输出值创建的。之所以只考虑时间窗口 TW 中的传感器输出值，而不是传感器过去的全部输出值，因为时间太久的传感器输出值对预测未来某个时间的传感器输出值不具有参考价值，而且计算、存储等资源的限制也不允许使用传感器过去所有的输出值。对给定一个时间点 $t>t^1$，$P_{s,t^0,t^1}(t,\upsilon)$ 返回 $S(t)=\upsilon$ 的概率值，称 $t-t^1$ 为预测范围。

简单预测模型是根据式（3）来计算 $S(t)=\upsilon$ 的概率值

$$p_{s,t^0,t^1}(t,\upsilon)=\frac{1}{t^1-t^0}\int_{t^0}^{t^1}\chi(s,q,\upsilon)d_q \tag{3}$$

式（4）是一个指示函数，如果传感器 S 在时间 t 的输出值是 υ，那么函数值等于 1，否则为 0。

$$\chi(s,t,\upsilon)=\begin{cases}1, & s(t)=\upsilon\\ 0, & \text{其他}\end{cases} \tag{4}$$

例如，如果传感器在过去的时间窗口 TW 中的输出值全部都是 υ，那么根据以上预测模型计算出的 $S(t)=\upsilon$ 的概率是 1，称上面的简单预测模型为聚集预测模型（APM）。

式（5）为单周期预测模型（SPPM），式（6）为多周期预测模型（MPPM）。

$$p_{s,t^0,t^1}(t,\upsilon)=\frac{1}{N}\sum_{1\leqslant i\leqslant N}\chi(s,t-iL,\upsilon) \tag{5}$$

$$p_{s,t^0,t^1}(t,\upsilon)=\max_{1\leqslant i\leqslant N}p_i \tag{6}$$

4.5.3　Dyser 总结

Dyser 系统把物理实体和传感器抽象为 Web 页面，方便了对物理实体的搜索，实现了和传统搜索引擎的无缝整合。其预测机制的使用不仅提高了搜索效率而且也降低了搜索开销，使其能够适用于资源受限的物联网环境。该系统的不足之处在于，预测机制的使用虽然提高了搜索效率但也降低了搜索结果的准确率。此外，Dyser 无法自动发现新加入的物理实体。

4.6　SenseWeb

由微软研究院设计的 SenseWeb[82]提供了基于静态元数据和基于位置的传感器搜索。SenseWeb 的系统架构如图 7 所示。

该系统的主要组成部分：协调器、传感器、传感器网关、移动代理、数据转化器和应用程序。传感器数据被传送到传感器网关，传感器网关提供了一个基于 SOAP 协议的 API，通过此 API 可以检索传感器信息和数据。所有的传感器网关都需要在协调器中进行注册，协调器是系统的核心。协调器包含任务和感知数据库 2 个模块。任务模块接受和分析来自应用程序的感知需求，为此需要考虑传感器的能力，共享意愿和其他特性。感知数据库缓存了

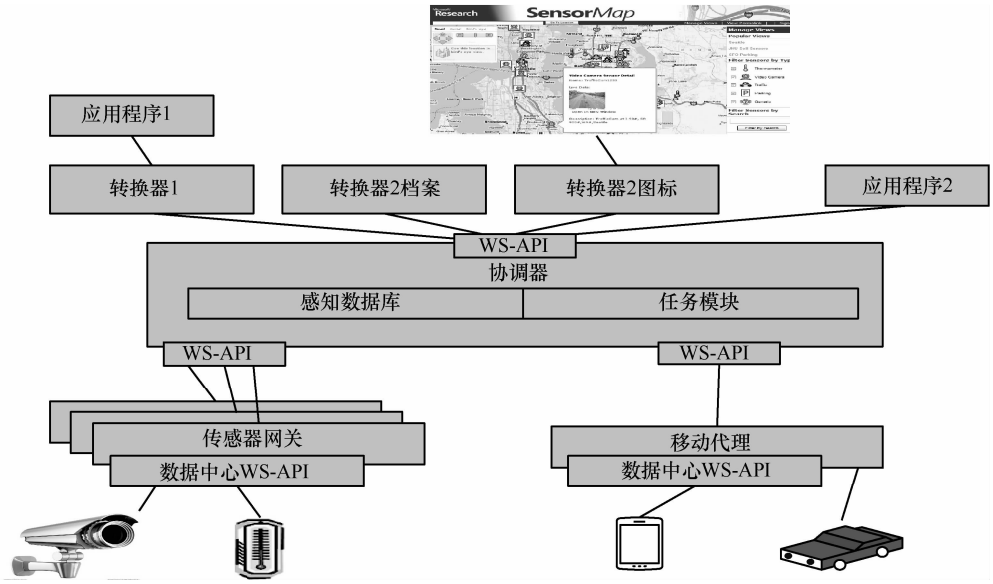

图 7 SenseWeb 系统架构

来自传感器网关的数据，可供多个应用程序访问，减少了对传感器数据的加载次数，节省了通信和存储资源。由于传感器和他们的位置注册在一个中心库中，传感器的移动会引起中心库中信息的更新，从而导致显著的开销。然而，系统引入了移动代理的概念。移动代理代表了一个固定位置的虚拟传感器。移动代理动态地绑定进入到其范围内的传感器。通过这种方式，中心库不需要直接支持传感器的移动性，解决了信息更新的问题，减少了开销。

在该系统中仅存储了传感器的静态元数据，因此，目前该系统仅支持基于元数据的静态搜索，不支持基于传感器内容的动态搜索。

4.7 DIS（distributed image search）

美国马萨诸塞大学安默斯特分校的研究人员设计的 DIS[83]是一个基于视频传感器网络的分布式图像搜索服务，该服务的每个节点都是一个搜索引擎，它可以感知、存储和搜索相关图像。DIS 的系统架构如图 8 所示。为了避免图像的传输，DIS 将图像转为特征向量进行存储和搜索，并为传感器平台设计一个基于闪存的查询优化的词汇树索引结构，用于图像搜索。DIS 可以帮助用户查询系统中已存储的图像，以及动态实时查询新捕捉的图像并及时通知查询结果，是一个针对图像的实时搜索引擎。搜索到的结果按照相似度从大到小进行排列，并且只返回相似度最大的前 k 个图。由于搜索请求被推送到所有的传感器，因此，DIS 无法扩展到大规模全球化的物联网系统。

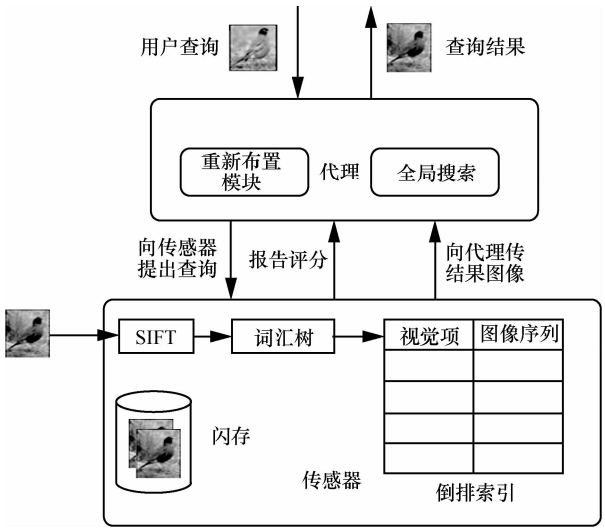

图 8 DIS 系统架构

4.8 RTS (real-time search)

RTS 是 Web 实时搜索引擎，RTS 搜索对象的内容是动态变化的。近年来由于社交网[84]的发展，Web 实时搜索引擎得到了快速发展。尽管 Web 实时搜索引擎不直接支持物理实体的搜索，但它的潜在机制能够被利用以实现针对物理实体的实时搜索。下面是 3 个 Web 实时搜索引擎的例子以及相关概念。

Twitter（推特）是当前非常流行的一个网站，它提供实时的基于 Web 的公开消息，也称为微博服务。Twitter 提供了一个网络平台，在这个网络平台中用户可以用文字（不超过 140 个字符）的方式实

时发布当前正在进行的活动。用户发送的消息被推送到 Twitter 服务器进行存档，Twitter 利用 XMPP 协议提供了一个可实时更新的公共消息池。推特搜索（search.twitter.com）是推特自己的搜索引擎。用户可以使用关键字在大量的 Twitter 消息中实时地搜索出最新的消息。

OneRiot (www.oneriot.com)是一个实时搜索引擎，重点通过社交网站如 Digg 和 Twitter 的用户共享链接。OneRiot 仅对被共享的网站进行索引，重点搜索与社交网站用户有关的内容。

Technorati.com 是一个微博搜索引擎。通过这个网站，Technorati 实时地检索数以百万计的博客文章，并能在几秒之内返回结果给用户。当用户更新微博时，通过调用一个专门的 API（叫做 remote procedure call）向搜索引擎发出一个提示消息。然而，最近该网站指出不再使用这种提示，因为多达 90%的这类提示信息是垃圾邮件或者非博客信息。由于物联网的大规模和高度动态性，这一方法并不适合于物联网。

搜索空间的受限性是上述实例的共同之处。Twitter 不仅严格限制了消息的大小，还限制了发布消息的速率。OneRiot 仅在社交网络空间中进行搜索。Technorati 仅考虑微博这一网络空间。它们的另外一个共同点是使用用户指定的提示信息，这影响了网站被重新索引的频率和搜索结果的排序。最后，Twitter 使用了集中式的方式把数据存储在 Twitter 服务器中，因此 Twitter 能够实时地查看用户发布的消息。

相比于人类产生的信息量，传感器产生的内容无论是在数量和更新频率上都要大很多。因此，物联网搜索面临的挑战要比以上 3 个实例复杂得多，然而，如何借鉴上述 3 个实例的关键技术实现一个物联网的实时搜索，这有待于进一步的研究。

4.9　其他系统

CASSARAM[85,86]提出了利用云计算解决物联网搜索的架构问题。该系统在物联网基础设施和服务的基础上建立了一个感知服务的云，通过云计算方式对传感器和传感器数据进行管理。CASSARM 还提出了基于情景感知的搜索方式以提高搜索的效率，涉及如用户偏好、传感器可靠性、精度、位置、电池寿命等情景信息；同时，CASSARAM 运用语义技术和定量推理提高了系统的性能。

文献[87]提出了一个基于时空、状态值、关键字的混合式物联网搜索引擎框架 IOT-SVK，它是一个支持多模态检索条件的搜索系统。传感器的位置信息可能随着时间的变化而改变，IOT-SVK 提出了用大小相等的网格区域代替原始的曲线路径，这解决了由于位置的动态变化造成索引的频繁更新而带来的开销问题。

文献[88]指出对于连续变化的传感器数据，搜索一段时间内的传感器数据值往往比搜索某一个时间点的传感器数据值更具有实际意义。该文使用模糊集的方法对搜索要求和传感器数据进行相似度匹配，以获得所要的搜索结果。该方法的使用节省了通信开销，解决了物联网搜索的大规模和实时性的问题。

文献[89]利用模糊集有效地比较传感器间输出值的相似性，从而实现传感器相似性搜索。文献[90]基于基于椭圆曲线设计了一个密码安全协议（ECC）的安全增强物联网搜索引擎。文献[91]根据智能设备的逻辑位置建立结构化的树形搜索图，位置属性是智能设备的主要属性，根据位置属性进行智能设备的搜索，并且提出了 4 种查询类型：详尽查询（EXQ）、基数查询（CAQ_k）、最佳努力查询（BEQ）、查询请求（RFQ）。表 2 给出了 8 个典型的物联网搜索系统之间的比较。

5　面临的挑战及展望

综上所述，物联网搜索的研究还处于起步阶段，对其关键技术还有待进一步的深入研究。目前已有的一些原型系统也只是针对某一个方面给出的解决方案，还无法应用于真正意义上的全球化的物联网，而且这些方案还未得到业界的一致认同和实际的广泛应用。因此，物联网搜索面临的挑战以及未来的研究空间都很大。

5.1　物联网搜索面临的挑战

1) 物联网搜索的架构设计。物联网搜索与传统的互联网搜索的不同，物联网搜索需要设计一种新的架构。然而，为物联网搜索设计一种合适的搜索引擎架构并非是件容易的事。因为从数据采集、爬取、索引、存储、用户搜索意图的理解和搜索方案的设计、知识挖掘到查询每个过程都不同于传统的搜索引擎，都需要一种新的技术。

2) 数据的有效表示和感知。物联网搜索对象广泛，包括网页、文档、音频、视频以及种类繁多的传感器数据。因此，物联网数据的有效感知和表示

表 2　　物联网搜索系统的比较

系统维度	Snoogle/Microsearch	MAX	OCH	GSN	SenseWeb	DIS	RTS	Dyser
体系结构	两层 分布式	三层 集中式	两层 分布式	基于容器的 集中式	两层 集中式	两层 分布式	文件系统	两层 集中式
聚合类型	混合	信标	定时器	信标	混合	混合	无	混合
索引类型	倒排索引	无	无	MySQL-style	倒排索引	倒排索引	倒排索引	倒排索引
查询类型	即席查询	即席查询	连续查询	即席查询	即席查询	连续查询和即席查询	即席查询	即席查询
查询方式	关键字	关键字	关键字	关键字	关键字和地理位置	图像	关键字	关键字
查询范围	本地	本地	本地	全局	全局	本地	全局	全局
查询时效	实时	实时	实时	实时	实时	实时和历史	实时（近似）	实时
查询精度	启发式	启发式	启发式	精确	精确	启发式	精确	精确
查询内容	静态	静态	动态	静态	静态	动态	动态	动态
查询结果	全部	全部	全部	全部	全部	Top-k	全部	Top-k
移动性	是	否	是	是	是	是	是	是
目标用户	终端用户	终端用户	终端用户	专家	专家	终端用户	终端用户	终端用户
安全支持	是	否	否	否	否	是	否	是

成为了一大挑战。

3) 物理实体的定位。位置信息是物联网实体中最重要的信息之一。如何在异构、多变的网络环境下实现物理实体的定位是物联网搜索面临的挑战之一。

4) 实时性。物联网最重要的功能之一是实时跟踪，然而由于传感器状态的高度动态性使传感器读数很容易过期，这给实时跟踪传感器读数，为用户提供准确实时的传感器数据带来了挑战。

5) 大规模性和异构性。物联网最大的特征之一是数据的大规模和异构性。物联网的大规模和异构性给通信资源、通信方式、数据存储、数据融合、数据检索等带来了挑战。

6) 语境与语义理解[92]。传感器捕捉到的是原始数据。然而，用户感兴趣的是物理现象而不是传感器的原始数据。例如，人们会搜索一个安静的地方，而不是搜索音量小于 30 dB 的地方。因此，需要相关机制把原始数据解释成物理现象。物联网搜索应能够自动、智能化地识别物理实体的语义信息；同时，结合用户请求的上下文、用户的情绪及历史偏好、被搜索对象所在环境的情景信息、时空特性等因素智能地分析用户的搜索意图，引导用户制定更好的搜索方案。

7) 安全和隐私问题。物联网搜索的安全问题以及隐私问题比传统的互联网更加重要和复杂。对于网页，人们可以选择不使用，或者把其设置为不可访问。但是，当传感器嵌入到每一个物理实体（如一件衣服）时，用户可能都不知道它们的存在。此外，传感器的资源受限，在设计安全协议时必须考虑资源消耗的问题，这使在物联网上实施安全管理更加困难。

5.2　物联网搜索研究展望

1) 物联网搜索的架构问题

未来物联网搜索系统的设计原则应该是分布式的、并行的、松散耦合的。未来的物联网搜索系统是由许多分布式的子搜索服务构成的，它们之间相互协作共同完成搜索任务。如何设计物联网搜索系统的子搜索服务的架构以及子服务之间如何协作都是需要进一步研究的问题。

2) 针对物联网高度动态性的解决方案

物联网实体高频率的动态加入和退出以及物理实体状态信息的高度动态性是物联网搜索面临的关键难点问题，这使传统的网络发现机制以及索引机制不再使用。现有的系统已经提出了预测机制和按时间段搜索 2 种解决方法。预测是指通过对历史数据的统计分析挖掘出规律，从而利用挖掘出的规律预测未来某个时间物理实体的状态信息。Dyser 用卷积和傅里叶变化的方式预测有周期规律的物理实体的周期，这种预测方式针对的是有周期规律的物理实体，显然无法应用于其他物理实体。现有

的预测机制还不完善和准确，如何利用预测理论方面的知识（如机器学习、灰色理论预测等）设计出较完善和准确的预测机制这是未来的一个研究方向。除了预测机制和按时间段搜索这 2 种方案，提出别的解决实体高度动态性问题的解决方法也是值得研究的问题。

3) 基于情景感知和推理实现物联网的高效和智能化搜索

对用户和物联网实体情景信息的感知、推理和理解能够提高物联网搜索的效率和准确性。情景信息有显式和隐式 2 种。物联网搜索应该具有感知和提取显式的情景信息能力，并且根据推理机制推出隐式的情景信息。

4) 物联网搜索语言的研究

物联网搜索应该支持多模态条件的检索，例如用户可以同时输入传感器类型、传感器位置、传感器状态值、时间值等多模态条件进行检索。然而，目前大多搜索都是采用关键字形式的方式进行搜索，基于关键字形式的搜索无法准确、全面地表达用户的搜索请求。因此，需要一种描述能力强的搜索语言提高搜索服务的性能和体验度。

5) 物联网中数据融合和数据挖掘的研究

物联网的搜索数据具有强烈异构性、多元性、多模态性、多属性和多维度性特征，为了保证搜索的质量，需要在物联网搜索中对获取到的各类搜索数据进行深度分析与融合，才能得到准确的搜索结果。由多个传感器融合后的信息可以更精确、更全面、更可靠，这是单个传感器无法完成的。对物联网数据进行数据挖掘可以挖掘出隐藏的更抽象的有价值的信息。

6) 物联网搜索与传统互联网搜索引擎的无缝整合

将物联网技术与成熟的 Web 技术相结合，把物理实体用传统网页的形式表示，对物理实体的搜索同时也转化成了对网页的搜索是物联网搜索的一个研究方向。另外，现有的互联网搜索引擎（如 Google，Baidu，Bing 等）已经表现得相当完善和成熟，并且已设计出大量先进的算法和技术。所以将物联网搜索和现有的搜索引擎进行无缝整合也是物联网搜索的一个研究方向。

7) 物理网搜索的安全问题

物联网搜索范围广，即包含传统的网页、文档、视频、音频，又包含物理世界中的实体。因此物联网搜索的安全隐私问题比传统的搜索引擎更加重要和复杂，既要面临传统信息安全的各种问题，又要面对物联网自身的特殊问题，如物理实体的访问控制和隐私性等。物联网搜索的安全问题也是未来的一个研究方向。

6　结束语

物联网搜索的对象由传统的网页、文档等延伸到物理世界，物理世界信息量大且复杂多变，这导致物联网搜索比传统的互联网搜索面临更多更大的挑战。本文分析了物联网搜索的相关概念及其框架，物联网搜索的特点，物联网搜索的相关技术，对现有的系统和算法进行了比较总结，最后分析了目前研究中存在的问题，并展望了其未来的发展方向。

参考文献：

[1] ATZORI L, IERA A, MORABITO G. The Internet of Things: a survey[J]. Computer Networks, 2010, 54(15): 2787-2805

[2] LIU H, BOLIC M, NAYAKAND A, et al. Taxonomy and challenges of the integration of RFID and wireless sensor networks[J]. IEEE Network, 2008, 22(6): 26-35.

[3] ENGLUND C, WALLIN H. RFID in wireless sensor network, EX034/2004[R]. Sweden: Communication Systems Group, Department of Signals and Systems, Chalmers University of Technology, 2004.

[4] 毛伟. 互联网资源标识和寻址技术研究[D]. 中国科学院计算技术研究所计算机系统结构,2006.
MAO W. Research on Internet Resource Identification and Addressing Technology[D]. Institute of Computing Technology Chinese Academy of Sciences, 2006.

[5] BRIN S, PAGE L The anatomy of a large-scale hypertextual Web search engine[J]. Computer Networks & ISDN Systems, 1998, 30(98): 107-117.

[6] DING C H, BUYYA R. Guided Google: a meta search engine and its implementation using the Google distributed Web services[J]. International Journal of Computers & Application, 2004, 10(3).

[7] 宋春阳, 金可音. Web 搜索引擎技术综述[J]. 现代计算机（专业版）, 2008, (05): 82-85.
SONG C Y, JIN K Y. Review of Web search engine technology[J]. Modern Computer (Professional Edition), 2008(05): 82-85.

[8] INMON B, Structured and unstructured data[EB/OL]. http://www.b-eye-network.com/view/4955. 2010.

[9] TAN C C, SHENG B, WANG H D, et al. Microsearch-when search engines meet small devices[A]. Proc of the 6th International Conference on Pervasive Computing[C]. Springer,2008:93-110.

[10] DAS R, HARROP P. RFID forecasts, players and opportunities 201-2021[R]. IDTechEx.com, 2010.

[11] ROMER K, OSTERMAIER B, MATTERN F, et al. Real-time search for real-world entities: a survey[J]. Proceedings of the IEEE,2010, 98(11): 1887-1902.

[12] 王智, 潘强, 邢涛. 面向物联网的实体实时搜索服务综述[J].计算

机应用研究,2011,28(6):2001-2010
WANG Z, PAN Q, XING T. Survey on real-time search service for entities of Internet of Things[J]. Application Research of Computers, 2011,28(6):2001-2010.

[13] 胡永利, 孙艳丰, 尹宝才. 物联网信息感知与交互技术[J]. 计算机学报, 2012, 35(6): 1147-1163.
HU Y L,SUN Y F,YIN B C. Information sensing and interaction technology in Internet of Things[J].Chinese Journal of Computers, 2012, 35(6):1147-1163

[14] SCHILIT B N, ADAMS N, WANT R. Context aware computing applications[A]. WMCSA[C]. Santa Cruz, CA, USA,1994. 85-90.

[15] DEY A K. Providing Architectural Support for Building Context Aware Applications[D]. Atlanta, GA, USA, Georgia Institute of Technology,2000.

[16] PERERA C, ZASLAVSKY A, CHRISTEN P, et al. Context aware computing for the Internet of Things: a survey[J]. IEEE Communications Surverys & Tutorials 2014,16(1):414-454.

[17] LANE N, MILUZZO E, LU H, et al. A survey of mobile phone sensing[J]. Communications Magazine, IEEE, 2010 48(9): 140-150.

[18] RIBONI D, BETTINI C. Context-aware activity recognition through a combination of ontological and statistical reasoning[A]. Proceedings of the 6th International Conference on Ubiquitous Intelligence and Computing, ser UIC '09[C]. Berlin, Heidelberg: Springer-Verlag, 2009. 39-53.

[19] w3.org. Swrl: a semantic web rule language combining owl and ruleml[EB/OL]. http://www.w3.org/Submission/SWRL/ 2012-01-03.

[20] ZHOU X, TANG X, YUAN X, et al. Spbca: semantic patternbased context-aware middleware[J]. IEEE International Conference on Parallel and Distributed Systems, 2009(12): 891-895.

[21] KESSLER C, RAUBAL M, WOSNIOK C. Semantic rules for context-aware geographical information retrieval[A]. Proceedings of the 4th European Conference on Smart Sensing and Context, ser EuroSSC'09[C]. Berlin, Heidelberg: Springer-Verlag, 2009. 77-92.

[22] CHOI C, PARK I, HYUN S, et al. Mire: a minimal rule engine for context-aware mobile devices[A]. Digital Information Management, Third International Conference on[C]. 2008, (10): 172-177.

[23] KONSTANTINOU N, SOLIDAKIS E, ZOI S A, et al. Priamos: a middleware architecture for real-time semantic annotation of context features[A]. Intelligent Environments, IET International Conference on[C]. 2007(9): 96-103.

[24] MANTYJARVI J, SEPPANEN T. Adapting applications in mobile terminals using fuzzy context information[A]. Proceedings of the 4th International Symposium on Mobile Human-Computer Interaction, ser Mobile HCI '02[C]. London, UK, UK: Springer-Verlag, 2002. 95-107.

[25] PADOVITZ A, LOKE S W, ZASLAVSKY A. The ecora framework: a hybrid architecture for context-oriented pervasive computing[J]. Pervasive Mob Comput 2008 4(2): 182-215.

[26] W3.org, Resource description framework (rdf)[S]. 2004.

[27] W3.org, Web ontology language (owl)[S]. 2007.

[28] RIBONI D, BETTINI C. Context-aware activity recognition through a combination of ontological and statistical reasoning[A]. Proceedings of the 6th International Conference on Ubiquitous Intelligence and-Computing[C]. Berlin, Heidelberg: Springer-Verlag, 2009. 39-53.

[29] TEYMOURIAN K, STREIBEL O, PASCHKE A, et al. Towards semantic event-driven systems[A]. Proceedings of the 3rd international Conference on New Technologies, Mobility and Security, Ser NTMS'09[C]. Piscataway, NJ, USA: IEEE Press, 2009. 347-352.

[30] NIST′s dictionary of algorithms and data structures: inverted index[EB/OL]. http://xw2k.nist.gov/dads//HTML/invertedIndex.html. 2008.

[31] ILYAS I,BESKAILES G, SOLIMAN M. A survey of top-*k* query processing techniques in relational database systems[J]. ACM Computing Surveys,2008 40(4):1131-1158.

[32] SMYTH B, BALFE E, BRIGGS P, et a1. Collaborative Web search[A]. Proceedings of the 18th International Joint Conference on Artficial Intelligence(IJCAI--03)[C]. Acapulco, Mexico, 2003. 1417-1419.

[33] SMYTH B, BALFE E, BOYDELI O, et a1. A live-user evaluation of collaborative Web search[A]. Proceedings of the 19th International Joint Conference on Artificial Intelligence(IJCAI-05)[C]. 2005. 1419-1424.

[34] BALFE E, SMYTH B. An analysis of query similarity in collaborative web search[J]. Lecture Notes in Computer Science, 2005 34(8): 330-344.

[35] MORRIS M R, HORVITZ E. Searchtogether: an interface for collaborative Web search[A]. Proceedings of the 20th ACM UIST Conference[C]. New York: ACM Press,2007. 3-12

[36] 周水庚, 李丰, 陶宇飞. 面向数据库应用的隐私保护研究综述[J]. 计算机学报, 2009 32(5): 847-861.
ZHOU S G,LI F,TAO Y F. Privacy preservation in database applications: a survey[J]. Chinese Journal of Computers,2009,32(5): 847-861.

[37] YAO A C. How to generate and exchange secrets[A]. Proc of the 27th IEEE Sym on Foundations of Computer Science (FOCS)[C]. Toronto, Canada, 1986. 162-167.

[38] CLIFTON C, KANTARCIOGLOU M, LIN X, et al. Tools for privacy preserving distributed data mining[J]. ACM SIGKDD Explorations, 2002, 4(2): 28-34.

[39] SWEENEY L. *K*-anonymity, a model for protecting privacy[J]. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002, 10(5):557-570.

[40] SWEENEY L. Achieving *K*-anonymity privacy protection using generalization and suppression[J]. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002, 10(5):571-588.

[41] CHOR B, GOLDREICH O, KUSHILEVITZ E, et al. Private Information Retrieval[J]. Journal of the ACM, 1998, 45(6): 965-982.

[42] KUSHILEVITZ E, OSTROVSKY R, Replication is not needed: single database, computationally-private information retrieval[A]. Proceedings of the 38th Annual Symposium on Foundations of Computer Science[C]. Florida, USA, 1997. 364-373.

[43] AGRAWAL R, EVFIMIEVSKI A, SRIKANT R. Information sharing Across Private Databases[A]. Proc of SIGMOD[C]. 2003,86-97.

[44] JARECKI S, LIU X. Efficient oblivious pseudorandom function with applications to adaptive ot and secure computation of set intersection[A]. TCC 2009[C]. LNCS,2009. 577-594.

[45] FREEDMAN M J, NISSIM K, PINKAS B. Efficient private matching and set intersection[A]. EUROCRYPT 2004[C]. Springer-Verlag (LNCS 3027), 2004. 1-19.

[46] STEFANOV E, SHI E, SONG D. Policy-enhanced private set intersection: sharing information while enforcing privacy policies[A]. PKC[C]. 2012. 203-245.

[47] ROY S, CONTI M, SETIA S, et al. Secure data aggregation in wireless sensor networks[J].IEEE Transactions on Information Forensics and Security,2012,7(3):1040-1052.

[48] RIVEST R, ADLEMAN L, DERTOUZOS M. On Data Banks and

Privacy Homomorphism. Foundations of Secure Computation[M]. New York: Academic Press, 1978. 169-179.

[49] FERRER J D. A provably secure additive and multiplicative privacy homomorphism[A]. Proc of the 5th International Conference on Information Security[C]. London: Springer-Verlag Press, 2002. 471-483.

[50] CASTELLUCCIA C, MYKLETUN E, TSUDIK G. Efficient aggregation of encrypted data in wireless sensor networks[A]. Proc of the 2nd Conference on Mobile and Ubiquitous Systems[C]. Washington: IEEE Computer Society Press, 2005. 109-117.

[51] MYKLETUN E, GIRAO J, WESTHOFF D. Public key based crypto schemes for data concealment in wireless sensor networks[A]. Proc of IEEE International Conference on Communications[C]. New York: IEEE Communications Society Press, 2006. 2288-2295.

[52] WESTHOFF D, GIRAO J, ACHARYA M. Concealed data aggregation for reverse multicast traffic in sensor networks: encryption key distribution and routing adaptation[J]. IEEE Transaction on Mobile Computing, 2006, 5(10):1417-1431.

[53] MLAIH E, ALY S A. Secure hop-by-hop aggregation of end-to-end concealed data in wireless sensor networks[A]. Proc of conference on Compute Communications. Washington, IEEE Computer Society Press[C]. 2008. 1-6.

[54] RODHE I, ROHNER C. *n*-LDA: *n*-layers data aggregation in sensor networks[A]. Proc of the 28th International Conference on Distributed Computing Systems Workshops[C]. Beijing: IEEE Computer Society Press, 2008. 400-405.

[55] BAHI J, GUYEUX C, MAKHOUL A. Secure data aggregation in wireless sensor networks: homomorphism versus watermarking approach[A]. Proc of Conference on Ad Hoc Networks[C]. Canada: ADHOCNETS Press, 2010. 344-358.

[56] FENG TAIMING, WANG CHUANG, ZHANG WENSHENG. Confidentiality protection schemes for data aggregation in sensor networks[A]. Proc of IEEE International Conference on Communications[C]. 2008. 1-9.

[57] OZDEMIR S, CAM H. ESPDA: Energy efficient and secure pattern based data aggregation for wireless sensor networks[A]. Proc of the 2nd IEEE Conference on Sensors[C]. New York: IEEE Society Press, 2003.

[58] SANLI H, OZDEMIR S, CAM H. SRDA: Secure reference-based data aggregation protocol for wireless sensor networks[A]. Proc of the IEEE VTC fall conference[C]. Los Angeles, 2004. 4650-4654.

[59] HE W B, NAHRSTEDT K, NGUYEN H. PDA: Privacy-preserving data aggregation in wireless sensor networks[A]. Proc of 26th IEEE International Conference on Computer Communications[C]. Washington, IEEE Computer Society Press, 2007. 2045-2053.

[60] ZHANG WEN SHENG, WANG CHUANG, FENG TAIMING. GP2S: generic privacy-preserving solutions for approximate aggregation of sensor data[A]. Proc of the 6th Annual IEEE International Conference on Pervasive Computing and Communications[C]. Hong Kong, China, 2008. 179-184

[61] 黎为.无线传感器网络数据融合安全方案的研究[D].长沙:湖南大学, 2009.
LI W. The research on secure data aggregation schemes in wireless sensor networks[D].Changsha: Hunan University,2009.

[62] GROAT M M, HE W B, FORREST S. KIPDA: *k*-indistinguishable privacy-preserving data aggregation in wireless sensor networks[A]. Proc of the 30th IEEE International Conference on Computer Communications[C]. Shanghai, China, 2011. 2024-2032.

[63] LI H J, LIN K, LI K Q. Energy-efficient and high-accuracy secure data aggregation in wireless sensor networks[J]. Computer Communication, 2011, 34:591-597.

[64] 杨庚, 王安琪, 等. 一种低能耗的数据融合隐私保护算法[J]. 计算机学报, 2011, 34(5): 792-800.
YANG G, WANG A Q, et al. An energy-saving privacy-preserving data aggregation algorithm[J].Chinese Journal of Computers, 2011 34(5): 792-800.

[65] BISTA R, JO K J, CHANG J W. A new approach to secure aggregation of private data in wireless sensor networks[A]. Proc of the 8th IEEE International Conference on Dependable Autonomic and Secure Computing[C]. Chengdu, China, 2009. 394-399.

[66] BISTA R, KIM H D, CHANG J W. A new private data aggregation scheme for wireless sensor networks[A]. Proc of the 10th IEEE International Conference on Computer and Information Technology[C]. Bradford, UK, 2010. 273-280.

[67] BISTA R, YOO H K, CHANG J W. A new sensitive data aggregation scheme for protecting integrity in wireless sensor networks[A]. Proc of the 10th IEEE International Conference on Computer and Information Technology[C]. Bradford, UK, 2010. 2463-2470.

[68] DU W, DENG J, HAN Y S. A witness-based approach for data fusion assurance wireless sensor networks[A]. Proc of IEEE Global Telecommunication Conference[C]. Washington, IEEE Computer Society Press, 2003. 1435-1439.

[69] GAO F, ZHU W T. A dual-head cluster based secure aggregation scheme for sensor networks[A]. Proc of the conference on Network and Parallel Computing[C]. Washington, IEEE Computer Society Press, 2008. 103-110.

[70] OZDEMIR S. Secure and reliable data aggregation for wireless sensor networks[A]. Proc of the 4th international Conference on Ubiquitous Computing Systems[C]. 2007. 102-109.

[71] OZDEMIR S. Functional reputation based reliable data aggregation and transmission for wireless sensor networks[J]. Computer Communications, 2008. 3941-3953.

[72] VU H, MITTAL N, VENKATESAN S. THIS: threshold security for information aggregation in sensor networks[A]. Proc of the 4th International Conference on Information Technology[C]. Washington, IEEE Computer Society Press, 2007. 89-95.

[73] BOHLI J.-M, VERARDI D, PAPADIMITRATORS P. Resilient data aggregation for unattended WSNS[A]. Proc of the 36th IEEE Conference on Local Computer Networks[C]. 2011. 994-1002.

[74] MAHIMKAR A, RAPPAPORT T S. SecureDAV: a secure data aggregation and verification protocol for wireless sensor networks[A]. Proc of the 47th IEEE Global Telecommunications Conference[C]. Dallas, TX, 2004.

[75] YANG Y, WANG X, ZHU S, et al. SDAP: a secure hop-by-hop data aggregation protocol for sensor networks[J]. ACM Transactions on Information System Secure, 2008, 11(18):1-43.

[76] LI DEPENG, ZEYAR A, WILLIAMS JR. Efficient authentication scheme for data aggregation in smart grid with fault tolerance and fault diagnosis[A]. Proc of IEEE Global Telecommunication Conference[C]. 2012. 1-8.

[77] WANG H D, TAN C C, LI Q. Snoogle: a search engine for pervasive environments[J]. IEEE Trans on Parallel and Distributed Systems, 2010, 21(8):1188-1202.

[78] TAN C C, SHENG B, WANG H, et al. BMicrosearch: when search engines meet small devices[A]. Proc 6th Int Conf Pervasive Comput[C]. 2008. 93-110.

[79] YAP K K, SRINIVASAN V, MOTANI M, BMAX: Human-centric search of the physical world[A]. Proc 3rd Conf Embedded Netw Sensor Syst[C]. 2005. 166-179.

[80] FRANK C, BOLLIGER P, MATTERN F, et al. The sensor internet at work: locating everyday items using mobile phones[J]. Pervasive Mobile Comput, 2008 4(3): 421-447.

[81] ABERER K, HAUSWIRTH M, SALEHI A. Infrastructure for data processing in large-scale interconnected sensor networks[A]. Proc Int Conf Mobile Data Manage, Mannheim[C]. Germany, 2007. 198-205.

[82] KANSAL A, NATH S, LIU J, et al. BSenseWeb: an infrastructure for shared sensing[J]. IEEE Multimedia, 2007 14(4): 8-13.

[83] YAN T, GANESAN D, MANMATHA R. Distributed image search in camera sensor networks[A]. SenSys[C]. 2008. 155-168.

[84] CORLEY A M. Real-time search stumbles out of the gate[J]. IEEE Spectrum, 2010, (12).

[85] PERERA C,ZASLAVSKY A,LIU C H, Sensor search techniques for sensing as a service architecture for the internet of things[J]. IEEE Sensors Journal, 2014, 14(2).

[86] PERERA C,ZASLAVSKY A,CHRISTEN P, et al. Context-aware sensor search, selection and ranking model for Internet of Things middleware[A]. 2013 IEEE 14th International Conference on Mobile Data Management[C]. 2013. 314-322.

[87] DING Z, GAO X, GUO L M. A hybrid search engine framework for the Internet of Things based on spatial-temporal, value-based, and keyword-based conditions[A]. IEEE International Conference on Green Computing and Communications, Conference on Internet of Things, and Conference on Cyber, Physical and Social Computing[C]. 2012. 17-25.

[88] TRUONG C, R¨OMER K. Content-based sensor search for the Web of things[A]. Global Communications Conference[C].2013. 2654-2660.

[89] TRUONG C, Romer K, CHEN K, Fuzzy-based sensor search in the Web of things[A]. Internet of Things (IOT), 2012 3rd International Conference on[C]. 2012. 127-134

[90] QIAN X J,CHE X P, Security-enhanced search engine design in Internet of Things[J]. Journal of Universal Computer Science,2012, 18(9): 1218-1235.

[91] MAYER S,GUINARD D,TRIFA V. Searching in a Web-based infrastructure for smart things[A]. 2012 3rd International Conference on the Internet of Things (IOT)[C].2012. 119-126.

[92] ZHANG D Q, YANG L T, HUANG H Y, et al. Searching in Internet of Things: vision and challenges[A]. Ninth IEEE International Symposium on Parallel and Distributed Processing with Applications[C]. 2011. 201-206.

作者简介：

高云全（1981-），男，安徽当涂人，北京邮电大学博士生，主要研究方向为物联网、可信服务。

李小勇（1975-），男，甘肃天水人，北京邮电大学副教授、博士生导师，主要研究方向为分布式计算与可信服务、网络安全、物联网。

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。