

# Information Retrieval for IoT and WoT: State-of-the-Art, Taxonomy Framework, and Evolutionary Directions

Cristyan Manta-Caro<sup>id</sup>, Annalina Caputo, and Juan M. Fernández-Luna<sup>id</sup>

**Abstract**—The explosive growth of Internet of Things (IoT) and Web of Things (WoT) technologies, characterized by a vast diversity of devices and data formats, producing vast volumes of information at a high pace in real time necessitates a paradigm shift in information retrieval (IR) systems. Traditional IR struggles to navigate the dynamic landscapes of these interconnected environments. This work proposes a multidimensional taxonomy framework to bridge this critical gap. Our framework not only unifies existing classification approaches but also delves into the analysis of traditional IR subtasks, thereby establishing a cohesive foundation for future advancements in IR tailored to the evolving IoT/WoT landscape. We further contribute by identifying key challenges and posing open research questions, thus propelling the development of next-generation IR techniques specifically tailored to the intricate search demands of the evolving IoT and WoT cyber-world.

**Index Terms**—Information retrieval (IR), Internet of Things (IoT), search engines, survey, Web of Things (WoT).

## I. INTRODUCTION

THE Internet of Things (IoT) drives new trends and paradigms that enable close interaction between end-users and the real world. To promote universal adoption and overcome scalability issues, the World Wide Web Consortium (W3C) and research communities advocate the integration of Web technologies. The Web of Things (WoT) provides Web-like access to advanced services while allowing interaction and manipulation of physical objects through virtual representations known as Avatar Web [1] a.k.a Digital Twins. WoT dynamics differ from the traditional Web due to real-time human interaction with intelligent cyber environments. WoT abstracts a massive number of physical and virtual objects in the real world by generating a vast amount of information at a high pace. As the WoT paradigm evolved, abstraction layers became complex, and technologies merged to create the Semantic WoT, leading to the vision of the Wisdom WoT.

IoT and WoT-based applications are changing the way we consume information. Information retrieval (IR) is one of the

most valuable forms of WoT-based applications, providing cyber-search functionalities to boost the smart cities paradigm. Next-Generation IR is a broad discipline that adds modern theoretical tools, such as learning, causal inference analysis, and interactive decision-making. IR systems seek relevant information that satisfies a user's information need within extensive collections [2].

IoT Search Engine (IoTSE) and WoT Search Engine (WoTSE) have been identified as one of the top 10 research topics in the IoT spectrum [3]. IoTSE and WoTSE refer to systems that allow humans and machines to retrieve IoT content, such as sensory data and digital representations of physical entities. While the terms are sometimes interchangeable, IoTSE focuses on finding physical entities and searching for raw or linked data. In contrast, WoTSE can find services and actions performed by digital representations of things. WoTSE can also search for things based on social relationships and ideally provide universal search capabilities.

Traditional IR systems have yet to be adequately tested or evaluated in the constantly changing and dynamic environments of IoT and WoT, which present new risks and challenges. Therefore, the search community must reconsider the conventional IR systems, including their scope, architecture, and internal stages. In our survey on IoTSE/WoTSE, we aim to classify different aspects of this field comprehensively. We analyze the evolution of the IoT/WoT-IR process, characterizing the context of applying IR techniques to IoT and WoT for searching any type of information that fulfills the exact user needs in these environments. Our secondary goal is to identify open research topics in the short and long term; we suggest evolutionary directions for applying IR to IoT and WoT.

The main contributions of this article can be summarized as follows.

- 1) *Identification of Challenges in IoT/WoT for IR*: This work recognizes the limitations of traditional IR systems in the context of the IoT and WoT landscapes. It highlights the need for a paradigm shift in IR to accommodate the highly dynamic, interconnected nature of IoT/WoT data.
- 2) *Development of a Multidimensional Taxonomy Framework*: This work proposes a comprehensive, multidimensional taxonomy framework designed specifically for IR in IoT/WoT. This framework unifies

Received 5 August 2024; revised 13 November 2024; accepted 11 December 2024. (Corresponding author: Cristyan Manta-Caro.)

Cristyan Manta-Caro and Juan M. Fernández-Luna are with the Department of Computer Science and AI, University of Granada, 18014 Granada, Spain (e-mail: cristyanmanta@correo.ugr.es; jmfluna@decsai.ugr.es).

Annalina Caputo is with the School of Computing, Dublin City University, Dublin 9, D09 V209 Ireland (e-mail: annalina.caputo@dcu.ie).

Digital Object Identifier 10.1109/IJOT.2024.3522219

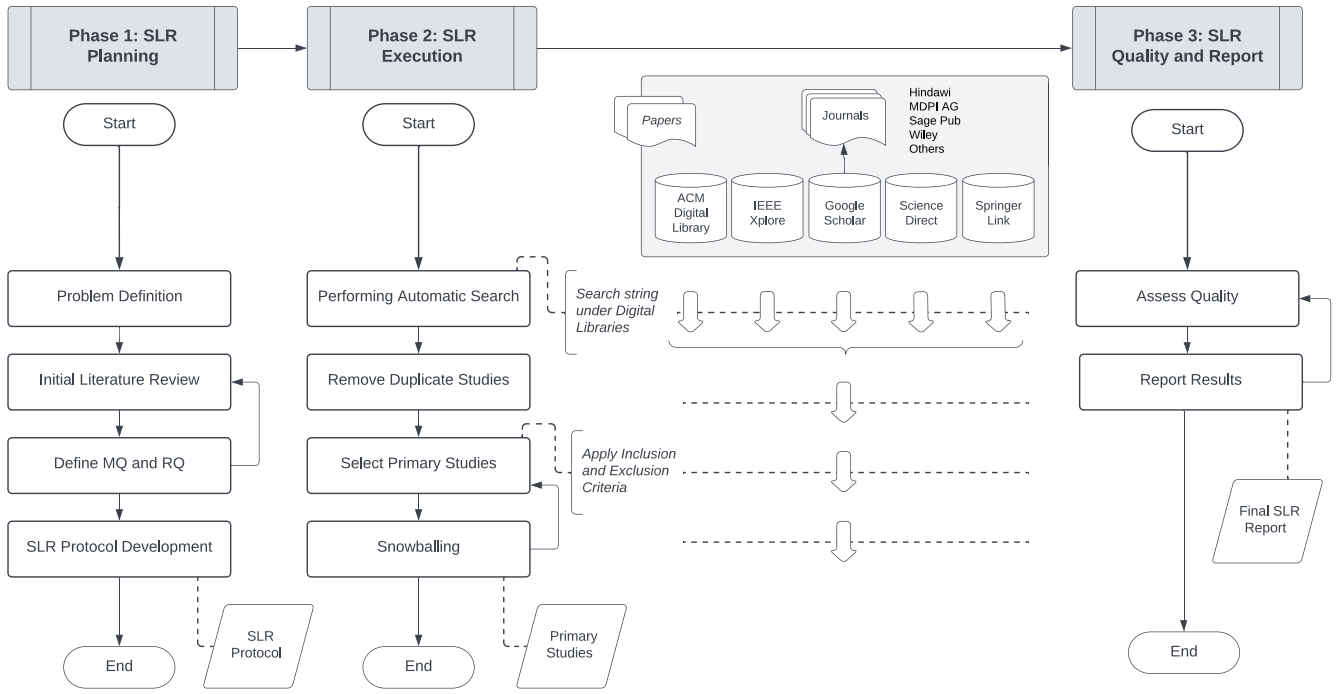


Fig. 1. SLR Protocol for IoT-IR and WoT-IR.

various classification methods already in use, providing a cohesive structure to analyze and categorize IR within these environments.

- 3) *Detailed Analysis of Traditional IR Subtasks:* The framework enables a deeper examination of traditional IR subtasks (such as indexing, query processing, ranking, and retrieval) and their adaptation to the requirements of IoT/WoT. The framework offers a robust foundation for future research.
- 4) *Identification of Key Challenges and Open Research Questions (RQs):* Besides the taxonomy, this article highlights specific challenges and unresolved questions within IR for IoT/WoT. These insights would encourage further research and innovation, paving the way for the development of next-generation IR systems.

This article structure is as follows. Section II describes the systematic literature review (SLR) performed, tracking the maturity of the pioneer works from 2000 to the present year 2024 related to IR's applicability to the IoT and WoT paradigms. Section III presents state-of-the-art details for each of the main IR subtasks. We contrast previous Survey-like works in Section IV, examining taxonomies and classification frameworks for IoTSE and WoTSE systems. This section also gives our proposed unified and holistic MultiTaxonomy Framework for IoTSE and WoTSE. In Section V, we explore the evaluation frameworks for IR-IoT/WoT. Section VI discusses open RQs, examining architectural and technical challenges on IR for IoT and WoT; we present short and long-term open research topics and evolutionary directions for the IR field, such as progressive search and IR on the social WoT. These insights can guide future research in the field. We

conclude with our final remarks, summarizing the essential findings and contributions of our work.

## II. SYSTEMATIC LITERATURE SELECTION, REVIEW, AND ANALYSIS METHODOLOGY

This article provides a comprehensive review of search engines for IoT and WoT; see Fig. 1. This section outlines the process performed for collecting and selecting primary studies, including the criteria for inclusion and exclusion. It delves into the IR process for IoT and WoT scenarios, defining RQs and selection strategies. Full details of the SLR can be found on GitHub and IEEE Data Port(TM). Therefore, this article is linked to SLR Protocol,<sup>1</sup> and the final version of the extracted data and analysis available as open dataset in GitHub.

*Definition of Research Questions:* We define the following mapping questions (MQs), which support and guide the scope of this SLR. *MQ1:* How many studies have been published over the years? *MQ2:* Who are the most active researchers in the area? *MQ3:* How has the IR subtasks on IoT and WoT evolved over the years? *MQ4:* Which challenges have been identified, and which open RQ still require new advancements?

To pursue the objectives of this work, we define the following RQs. *RQ1:* Is/Are there any other SLR and taxonomies developed for IR-IoT and IR-WoT? *RQ2:* Which IR subtasks have been used, impacted by or proposed for the IoT/WoT paradigms? *RQ3:* What are the basic foundations for IoT, WoT and the applicability of IR subtasks on those paradigms? *RQ4:* How the proposed solutions have been modelled and

<sup>1</sup>"SLR Protocol for IR for IoT and WoT"  
<https://github.com/cristyanmanta/IR-IoTSE-SLR-Protocol>.

algorithmically and technically structured to face IoT/WoT challenges?

*Definition of the SLR Scope (PICOC):* To help answer the RQs, we analysed the literature following the PICOC scope.

*Population (P):* “IoT” OR “Internet of Things” OR “WoT” OR “Web of Things.”

*Intervention (I):* “IR” OR “Search Engine” OR “Crawling” OR “Indexing” OR “Querying” OR “Retrieving” OR “Ranking” OR “Discovery” OR “Presenting.”

*Comparison (C):* Not Applicable.

*Outcomes (O):* Paper on the state-of-the-art, taxonomy framework and future directions for IR on IoT and WoT.

*Context (C):* Peer-reviewed and conference publications on IR subtasks, models, algorithms and techniques.

*Definition of the Inclusion (IC) and Exclusion Criteria (EC) (IC1):* Paper proposing at least one IR mechanism (subtask, method, technique, model, algorithm, search engine); AND *IC2:* The proposed solution is applied to IoT OR WoT OR precursor; AND *IC3:* This article publication period must be between 2000 and 2024. *IC4:* The study must be available in full text in selected digital libraries. *EC1:* The study is irrelevant to IoT/WoT or any of its precursor domains and the field of IR. *EC2:* The study is a similar or reduced version of a complete work. *EC3:* This article is a tertiary study or any nonscientific report. *EC4:* The study is written in a language other than English.

Our SLR approach uses ACM Digital Library, IEEE Xplore, Google Scholar, ScienceDirect, and Springer Link. The SLR process involves three phases.

- 1) In phase 1, the primary reviewer collects data from primary studies, bibliometrics, demographics, taxonomics (if applicable), features, and research info. The extracted data are tabulated for each study, and quality and self-assessment questions generate data for each paper.
- 2) In phase 2, our search strategy follows the methodology described in [4] and consists of four stages: a) an automatic search over the most relevant scientific digital libraries; b) removal of duplicate papers; c) consideration of only papers related to the topic following predetermined inclusion criteria; and d) further search by forward snowballing.
- 3) In phase 3, we assess the quality of the results and reports. We gather and structure the most outstanding results to analyse and discuss. Using a checklist, the researcher evaluates the aspects relevant to the SLR in each paper. Each paper is either included or excluded in the final phase for reporting, depending on the evaluation score.

The output of the SLR is a survey on IR systems for IoT and WoT. We review pioneer work and recent primary studies reported between 2000 and 2024 to provide a comprehensive analysis, see Figs. 2 and 3.

All in all, (4.3%) of works focus on pure data integration or fusion, and (23.4%) of works are oriented to discovering and crawling algorithms. We group them into one category because of their similarities. However, we shall highlight that both Search Scope/Space (3.3%) will advise using only one or

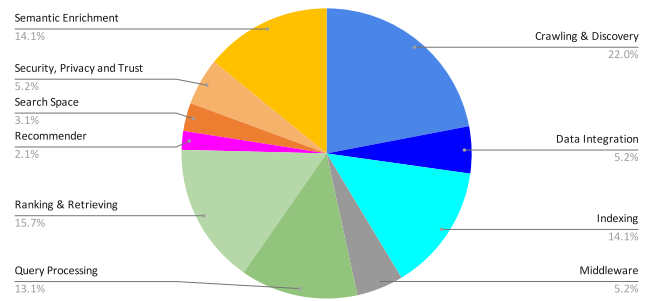


Fig. 2. Percentage of analysed studies by predominant IR subtask proposed or presented for IoT—WoT.

both types of algorithms. A minority set of approaches (14.7%) focuses on indexing mechanisms discussing data structures for indices or strategies. Middlewares deals with device heterogeneity and with communications protocol requirements in the form of translations or adaptations (what could be considered WoT Binding Templates in W3C terminology). (13.6%) of works have a predominant discussion on query processing mechanisms, the majority of them biased toward SPARQL-based proposals. (14.7%) of analyzed studies have proposed and presented ranking strategies or scoring methodologies for improving query performance or information relevancy. (13.6%) of research primarily aimed to provide semantic enrichment to “simplify” other mechanisms, add a knowledge base, link concepts, or enhance the abstraction models. Finally, we count (4.9%) of recent studies researching the security, privacy, and trust perspectives of search engines for IoT and WoT. Only a tiny fraction (2.2%) proposes recommender systems for IoT and WoT.

### III. STATE-OF-THE-ART IN SEARCHING IoT-WoT

#### A. Evolution and Pioneers of IR for IoT and WoT

Fig. 4 showcases the conceptual evolution of IoTSE/WoTSE and demonstrates the maturity achieved by the different branches grouped according to the leading research lines, including SPARQ/Non-IR approaches. Research on building search mechanisms for the predecessors of IoT/WoT systems dates back to the early 2000s, which we refer to as the precursors of IR-based IoTSE stage. The WoT concept has been in development since 2006. However, complete IoTSE/WoTSE systems were introduced at the end of the decade, exploring different aspects of the problem and contributing to the development of new perspectives in the IR field. Pioneering proposals, such as DYSER [5], SNOOGLE [6], and MICROSEARCH [7] were introduced during this time. These proposals helped pave the way for the advancement of the WoT concept and have significantly contributed to the development of the IR field. Fig. 4 shows the three primary areas of research: 1) searching for entities like people, places, and things using IR; 2) discovering and searching for resources and services, such as data, events, and IoT devices sensors/actuators without relying on IR; and 3) searching for data, streams, and linked data using SPARQL.

Our analysis focuses on advanced and innovative components of complete IR systems.

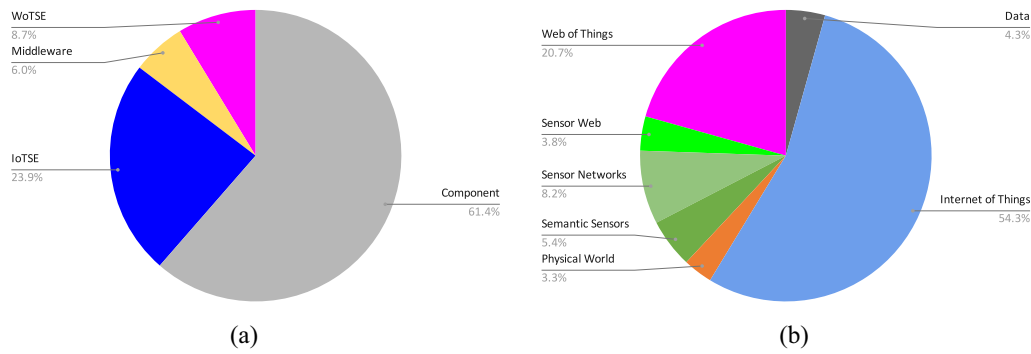


Fig. 3. Percentage of analysed studies (a) by proposal type: IoTSE, WoTSE complete systems or specific components and (b) by scenario.

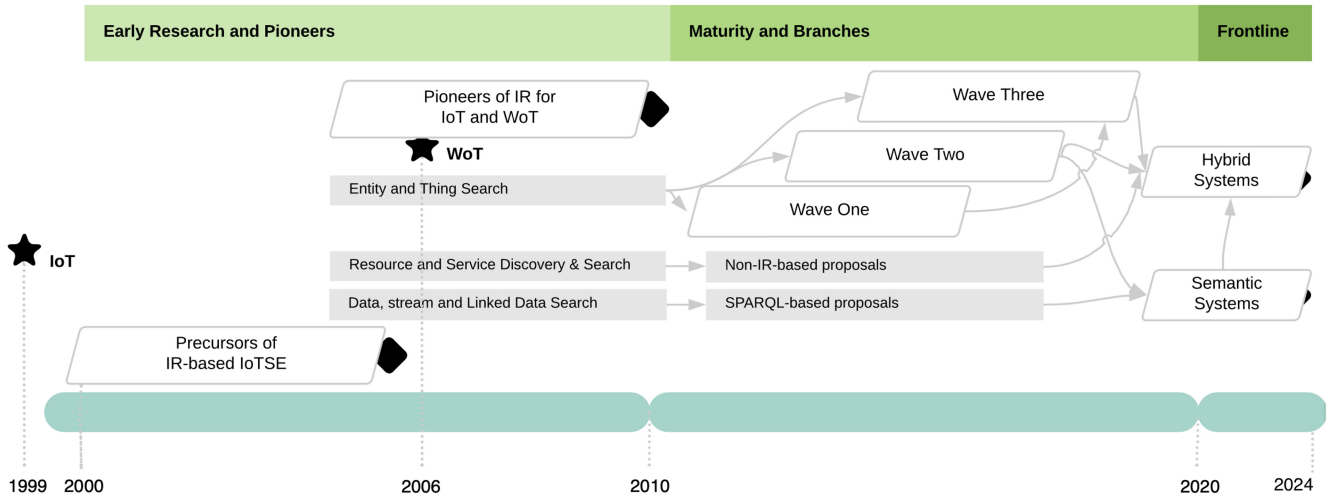


Fig. 4. Evolution of IR for IoT and WoT.

- 1) *Wave One*: We distinguish here those approaches that adapt some IR-like methods and techniques to extended or advanced versions of the predecessor systems of IoT/WoT. We include here standalone adaptations.
- 2) *Wave Two*: IR Pioneers for semantic WoT enrichment.
- 3) *Wave Three*: Multimodal & context-aware search, including distance-awareness.

Researchers on the front line are bringing new perspectives and cutting-edge technology to IR. They are incorporating artificial intelligence (AI) mechanisms into IoT and WoT architectures. For instance, Cheng et al. [8] introduced neural architecture search methodologies for IoT. These methodologies support a flexible search space and implement a progressive coarse-to-grained search mechanism. Although this study primary focus is not on IR, some new perspectives could be applied to new ranking or retrieving systems, especially for ranking results considering IoT metrics. It is important to note that, although the focus of this article is not on SPARQL and Non-IR approaches, these approaches were still considered when building the SLR dataset.

### B. IR Subtasks and Approaches for IoT/WoT

The discussion of a dynamic IR system for IoT/WoT has been divided into subtasks to present our findings and insights based on the utilization, impact, and proposals of IR subtasks for these paradigms. Key IoTSE/WoTSE-like works

published between 2002 and 2024 are listed in Table I. These findings will undoubtedly provide valuable insights into the development of IR systems for the IoT and WoT paradigms.

1) *Crawling and Discovering for IoT/WoT*: Crawling the WoT is still a proceeding area of research [59]. Classically, Crawler is a program that automatically scans websites by following links from one webpage to another. A WoT Crawler typically includes a Discovery mechanism and may be tailored to a specific domain. WoT Crawling should be distinguished from IoT Discovery. The architecture of the WoT Crawler must remain protocol-independent, considering the nuances of data exchange protocols used in IoT/WoT infrastructure. Nath et al. [56] defended the idea that the WoT Crawler might be application-specific to a domain and not generic. A WoT Crawler has three primary functions [28]: 1) identifying data sources; 2) finding and extracting metadata or semantic elements; and 3) integrating, linking, and correlating them to build an index system.

Discovery involves linking new data sources to existing systems through crawling algorithms [17]. Discoverers can be centralized or distributed, with a preference for decentralization; for example, recently, Chen et al. [12] proposed vector symbolic architecture (VSA)-SD: a distributed service discovery method for IoT devices based on VSA. VSA-SD uses hyperdimensional vectors to describe services and calculates Hamming distance for service discovery. Centralized

TABLE I  
 COMPILATION OF KEY IOTSE/WOTSE-LIKE WORKS BETWEEN 2002 AND 2024

Year	Authors	Reference	System Name	WoTSE — IoTSE	Data Type	Sensor Type	Thing Type	Context	IR Predominant Sub-Task
2024	Krishankumar et al.	[9]	—	Middleware	STD — DYD	PSN — VSN	PHT	CTX—SRV	Ranking and Retrieving
2024	Moghiss-Shameli-Sendi	[10]	I-RECON	IoTSE	STD — DYD	PSN — VSN	PHT	SEC	Security, Privacy and Trust
2024	Aguzzi et al.	[11]	ZION	WoTSE	STD — DYD	VSN	VTH — TD	CTX—W3C	Query Processing
2023	Chen et al.	[12]	VSA-SD	Middleware	DYD	PSN	ND	CTX—SRV	Crawling and Discovery
2023	Faheem et al.	[13]	—	WoTSE	STD — DYD	VSN	VTH — TD	CTX—W3C	Indexing
2022	Hatcher et al.	[14]	SecureIoT	IoTSE	STD — DYD	PSN	ND	SEC	Security, Privacy and Trust
2022	Mili-Rodin	[15]	WoTmAS2E	WoTSE	STD — DYD	PSN	PTH — AGT	CTX	Ranking and Retrieving
2022	Pavlopoulou-Curry	[16]	PoSSUM	Middleware	STD — DYD	ND	ND	ND	Semantic Enrichment
2021	Iggena et al.	[17]	IoTcrawler	IoTSE	STR — LDA	VSN	ND	ND	Crawling and Discovery
2021	Nandan et al.	[18]	IoT-CBSE	IoTSE	STD — DYD	PSN	PHT	ND	Semantic Enrichment
2020	Sciullo et al.	[19]	AHF-WoT	WoTSE	STD — DYD	VSN	VTH — TD	CTX	Crawling and Discovery
2020	Tang et al.	[20]	CESCE	IoTSE	STD — DYD	PSN	PHT	ND	Query Processing
2020	Yang et al.	[21]	PSSPP	IoTSE	DYD	PSN	ND	ND	Security, Privacy and Trust
2019	Pattar et al.	[22]	ProSA	WoTSE	DYD	PSN	PHT	CTX — SRV	Crawling and Discovery
2019	Sciullo et al.	[23]	WoT Store	WoTSE	STD — DYD	VSN	VTH — TD	CTX — APP	Crawling and Discovery
2019	Silva et al.	[24]	SWoTPAD	WoTSE	DYD	ND	ND	ND	Semantic Enrichment
2019	Tang et al.	[25]	SMPKR	IoTSE	STD — DYD	PSN	PHT	ND	Query Processing
2018	Karim et al.	[26]	DESERT	Component	STR	ND	ND	ND	Query Processing
2018	Ma-Liu	[27]	PROVINS	Component	DYD — STA — IMG	PSN	ND	ND	Search Space
2018	Skarmeta et al.	[28]	IoTcrawler	Component	DYD — STR	PSN	ND	ND	Crawling and Discovery
2017	Shemshadi et al.	[29]	ThinkSeek	WoTSE	DYD — STR	PSN	ND	ND	Crawling and Discovery
2016	Nunes et al.	[30]	VisIoT	Component	DYD	ND	ND	ND	Crawling and Discovery
2016	Qin et al.	[31]	CEIoT	Middleware	STD — DYD	PSN	PHT — AGT	CTX — SOC	Data Integration
2016	Shemshadi et al.	[32]	ThinkSeek	WoTSE	DYD — STR	PSN	ND	ND	Crawling and Discovery
2016	Younan et al.	[33]	WoTSF	WoTSE	STD — DYD	VSN	VTH	ND	Indexing
2015	Ebrahimi et al.	[34]	AntClust	IoTSE	STD — DYD	SSN	ND	ND	Search Space
2015	Gomes et al.	[35]	ForwardS-IoT	IoTSE	SAT	SSN	ND	CTX — SRV	Middleware
2015	Li et al.	[36]	EasiCrawl	Component	DYD — STR	PSN	ND	ND	Crawling and Discovery
2015	Soldatos et al.	[37]	OpenIoT	IoTSE	STD	SSN	ND	ND	Middleware
2014	Kolcun-McCann	[38]	Dragon	Component	STD	PSN	ND	ND	Query Processing
2014	Michel et al.	[39]	Gander	IoTSE	STD — DYD	PSN	PHT	ND	Query Processing
2014	Perera et al.	[40]	CASSARAM	IoTSE	STD — STR	SSN	PHT	ND	Ranking and Retrieving
2013	Jara et al.	[41]	MobDigcovery	Component	DYD	PSN	ND	ND	Crawling and Discovery
2012	Ding et al.	[42]	IoT-SVK	IoTSE	STD — DYD	ND	ND	ND	Indexing
2012	Jara et al.	[43]	ImDNS-SD	Component	DRC	VSN	PHT	ND	Crawling and Discovery
2012	Le-Phuoc et al.	[44]	LSM	Component	DYD — SAT	ND	ND	ND	Middleware
2012	Pintus et al.	[45]	Paraimpu	IoTSE	DYD	PSN	ND	ND	Recommender
2012	Qian-Che	[46]	ISE	IoTSE	STD	ND	ND	ND	Security, Privacy and Trust
2011	Komatsuzaki et al.	[47]	IteMinder	IoTSE	STD	ND	ND	ND	Search Space
2011	Le-Phuoc et al.	[48]	CQELS	Middleware	STR — LDA — SAT	ND	ND	ND	Middleware
2011	Mayer-Guinard	[49]	DiscoWoT	WoTSE	STD	ND	VTH	ND	Crawling and Discovery
2011	Pfisterer et al.	[50]	SPITFIRE	WoTSE	DYD — SAT	PSN — VSN	ND	ND	Semantic Enrichment
2010	Ostermaier et al.	[5]	Dyser	WoTSE	DYD	VSN	VTH	ND	Indexing
2010	Tan et al.	[7]	Microsearch	IoTSE	STD	ND	ND	ND	Crawling and Discovery
2010	Wang et al.	[6]	Snoogle	IoTSE	STD	PSN	PHT	ND	Crawling and Discovery
2009	Elahi et al.	[51]	SensorRanking	Component	STD — STA	VSN	VTH	ND	Ranking and Retrieving
2009	Henson et al.	[52]	SemSOS	IoTSE	SAT	VSN — SML	ND	ND	Semantic Enrichment
2009	Jirka et al.	[53]	OSIRIS	Component	DYD	VSN — SML	ND	ND	Crawling and Discovery
2008	Frank et al.	[54]	OCH	IoTSE	STD	ND	PHT	ND	Query Processing
2007	Kansal et al.	[55]	SenseWeb	Middleware	STD	SWB	ND	ND	Middleware
2007	Nath et al.	[56]	SensorMap	Component	DYD	PSN	ND	ND	Crawling and Discovery
2005	Yap et al.	[57]	MAX	WoTSE	STD	ND	PHT	CTX	Query Processing
2003	Li et al.	[58]	DIM	Component	DYD	PSN	ND	ND	Query Processing

Table I convention is as follows:

- Data Type: STD Static Data; DYD Dynamic Data; LDA Linked Data; IMG Image; STR Streaming Data; SAT Semantic Annotation; STA States.
- Sensor Type: PSN Physical Sensor; VSN Virtual Sensor; SWB Sensor Web; SML Semantic SensorML, ND Non-Described.
- Thing Type: PHT Physical Thing; VTH Virtual Thing; AGT Agent based or Avatar; TD Thing Description (W3C); ND Non-Described.
- Context or Additional Abstraction: SEC Security-awareness; CTX Context-Awareness; SRV Service-based; APP Application-based; SOC Social-awareness.

approaches build a registry of IoT resources and services, while distributed approaches may be based on location or infrastructure and use layered or clustered architecture. In traditional Web crawlers, search engines identify the existence of Web pages and continuously crawl for new and updated resources. Automatic crawl mechanisms for WoT can be likened to WoT crawling if the aim is to build a repository collection. It is crucial to consider the impact of having a central repository when assessing the scalability of any solution. We must highlight the major crawling proposals for WoT: IOTCRAWLER [17], [28], EASICRAWL [36] and THINGSEEK [32]. Alternatively, a simple but less scalable solution involves static registries where users manually enter functionalities or IoT services [60]. Future research will involve advancements in both crawler design and standardized

discovery mechanisms to bridge the gap between the WoT and the Web of Data.

2) *Indexing, Data Structures, and Strategies for IoT/WoT:* Indexing is a crucial subtask in IR systems, particularly for the WoT. An IR system performs two primary functions: 1) indexing and 2) query processing. Indexing involves creating efficient data structures for retrieving information from a collection of documents, mainly text. Query processing uses these index data structures to generate a ranked list of documents for a user's query. Many studies have explored the functions, benefits, and drawbacks of data structures in the context of IoT and WoT, including multidimensional approaches like R-tree, R\*-tree, SR-Tree, X-tree, kD-tree, VA-file, and Pyramid. In [61], the advantages and disadvantages of using data structures are presented.

IR solutions often rely on static indexes, which are not well-suited for the dynamic nature of the WoT. To accommodate this dynamism, dynamic indexing for WoT must be data-independent and scalable. Various dynamic indexing techniques have been proposed for both IoT and WoT, including different data structures and strategies.

Dynamic indexing for WoT requires data independence and affordable scalability. Tran et al. [2] classify the Index Type as Text-based, Spatial Indexes, numerical Value Indexes, Clustering mechanisms, Prediction Models, and Unspecified Indexes. We can identify different directions in the dynamic indexing techniques for IoT and WoT for both data structures and strategies; some proposals include:

- 1) index relies on a centralized mechanism. Traditional or specialized DB technologies, e.g., GeoDB, GraphDB;
- 2) index relies on a sort of Registry, Directory, or Catalogue;
- 3) index is implemented as an Inverted file similar to traditional Search Engines;
- 4) signature Indices based on hashing approaches; and
- 5) clustering Indices.

To achieve optimal efficiency and effectiveness, specialized indexing structures are necessary for different types of data. While tree and hashing models are common, R-tree is beneficial for geospatial data, R+/MDR+-tree for spatio-temporal data, and RtGR-tree for observation sensor data [62]. Other structures like PKR-tree and STK-tree have also been studied. To meet user needs beyond conventional text or spatial indexing, specialized index schemes have been proposed for spatio-temporal, thematic, and near-real-time information.

These indexes can be distributed, multiindex, or hierarchical. However, constructing and maintaining these indexes can be challenging, and few studies have focused on maintenance strategies. Some proposed similarity scores and clustering for thematic indexes, while others suggested strategies for maximizing freshness while keeping computational costs minimal [63]. Recent research has explored various indexing strategies for IoT streaming data. Doan et al. [64] introduced a framework focusing on query optimization, comparing B-Plus trees and hash tables while employing compression and summarization techniques for efficiency. Similar optimization approaches are presented in [65], utilizing dynamic time warping for data reduction. A categorization of indexing strategies into full, centralized, and distributed hash tables (DHTs) is proposed in [66]. Full indexing involves building a comprehensive data index table (DIT) across all IoT edge nodes, potentially impacting freshness and scalability. Centralized indexing addresses scalability but introduces a single point of failure concerns. DHT-based approaches distribute the index across nodes, improving fault tolerance but increasing query complexity. Dynamic indexing for IoT data is explored in [67], covering multidimensional and metric indexing techniques within cloud-fog computing environments. The authors discuss partitioning strategies for index distribution across edge nodes. In a different direction, Faheem et al. [13] proposed a machine learning-based approach, combining indexing, clustering, and semantic modeling to create a searchable database for indoor things, enhancing efficiency and accuracy in intelligent environments.

This overview highlights the diverse landscape of IoT data indexing research, encompassing optimization techniques, indexing structures, and deployment strategies.

3) *Querying, Ranking, and Retrieving for IoT/WoT*: Standardization has been a guiding principle for the proposals found in the literature. SPARQL and its extensions and derivatives play an important role in Semantic WoT by facilitating an integrated IoT/WoT ecosystem. However, there has yet to be an agreement on which ontologies should be utilised, and a mechanism is still needed to allow end-users to interact with WoT in a human-centric manner. While low-level queries are possible, natural language or another approach should be provided for high-level queries from the perspective of the human end-user. We suggest a straightforward categorization of query interpreters based on their query language and capabilities.

- 1) *Low-level* those able to provide a coherent and straightforward search mechanism for RDF ontology-based proposals at sensor and data layers. In this, we group SPARQL, its derivatives and extensions, RDF<sup>2</sup> data query language (RDQL), or adaptations, and semantic Web rule language (SWRL).<sup>3</sup>
- 2) *High-level* query interpreters and languages provide richness in the specification of user expectations and information needs. Some can receive the query in natural language, while others provide an alternative synthetic language. New approaches have been proposed during the maturity of IoTSE/WoTSE research. Du et al. [68] proposed an IoT-WS query as a tuple consisting of distance, time and functionality. It provides a simple mechanism, but in contrast, it can restrict the expressiveness and richness of the query itself.

Multiresolution queries, introduced in [25] and [69], expanded the scope of IoT/WoT search beyond keyword-based queries. Tang et al. proposed SMPKR for spatio-temporal keyword-based search using PKR-tree indexing. Building upon this, Tang et al. [70] introduced CECSE, a collaborative edge-cloud cache-based WoTSE for mobile objects. While these works advanced the field, challenges remain in integrating diverse data sources and mechanisms for comprehensive IoT/WoT search.

CECSE employed a three-tier cache architecture and SKIN-tree indexing for efficient query processing. Diamantini et al. [71] proposed a multiresolution, multi-granularity, context-based approach. Complementarily, Ma and Liu [27] focused on search progressiveness. A recent contribution, ZION, described in [11], is an open-source W3C Thing Description Directory that efficiently queries W3C Thing Descriptions, offering scalable CRUDL operations and JSONPath metadata search.

The richness of IoT/WoT data presents opportunities and challenges for search. While prediction models can enhance search accuracy, limitations in prediction accuracy and computational overhead persist. Zhang et al. [72] proposed a dual-mode sensor search mechanism and improved prediction

<sup>2</sup><https://www.w3.org/Submission/RDQL/>

<sup>3</sup><https://www.w3.org/Submission/SWRL/>

models. Liu et al. [73] explored combinatorial-oriented feedback for data sensor search. These approaches aim to reduce query processing overhead, optimize resource consumption, and improve search efficiency by focusing on relevant sensor data.

Overall, the field is evolving toward more sophisticated search capabilities, combining diverse data sources, advanced indexing techniques, and predictive models to address the complexities of IoT/WoT environments.

4) *Ranking and Retrieving in IoT—WoT*: Sensor ranking diverges significantly from traditional search engine relevance ranking, prioritizing search engine performance over user satisfaction. Perera et al. [40], [74] introduced semantic sensor search, enabling users to query based on sensor parameters, such as reliability, accuracy, location, and energy consumption. Their proposed comparative-priority weighted index (CPWI) ranks sensors by calculating a similarity score between user preferences and sensor attributes, employing techniques like fuzzy logic or weighted linear combination. Numerous subsequent studies have explored custom multicriteria scores incorporating additional parameters, such as sensor precision, latency, and cost-effectiveness.

Categorizing approaches by ordering and internal ranking methods reveals diverse techniques. DYSER [5] employs predictive models, such as time series analysis or machine learning, to forecast sensor performance and rank them accordingly. SNOOGLE [6] utilizes a *tf-idf* weighting scheme to assign relevance scores to IoT objects based on term frequency and inverse document frequency, similar to traditional text retrieval. Microsearch [7] also leverages *tf-idf* within a top-k retrieval framework to generate ranked lists of sensors. LIVEWEB [75] adopts a Boolean model, matching sensor attributes to query criteria using logical operators. Mili-Rodin's [15] WOTMAS2E introduces a probability-based ranking process, considering sensor states and their associated probabilities to determine rankings.

Despite these advancements, ranking remains a secondary focus in the literature. Fathy et al. [76] defined data ranking as prioritizing IoT resources based on Quality of Information (QoI) and Value of Information (VoI), incorporating metrics, such as accuracy, completeness, consistency, and relevance.

Recent work has expanded ranking methodologies. Parreira et al. [77] proposed a multicriteria ranking strategy using a weighted Quality of Experience (QoE) score, considering factors like data freshness, accuracy, completeness, and consistency. Truong et al. [78] introduced a fuzzy-based similarity score, calculating the degree of match between sensor attributes and query criteria. Additionally, Krishankumar et al. [9] applied fuzzy logic to IoT Service Provider (IoTSP) selection, considering factors like mobility, security, and connectivity, and ranking providers based on overall fuzzy membership values.

These approaches contribute to the evolving landscape of IoT/WoT service and resource ranking in IoTSE/WoTSE systems, with a growing emphasis on incorporating multiple criteria, utilizing advanced ranking algorithms, and considering user preferences and system performance.

5) *Presenting UI/UX for IoT/WoT*: Different strategies have been considered for presenting results in the world of WoTSE research. These strategies have different formats and scopes that can complement traditional IR systems facing end-users. The presentation layer has been evaluated based on user type, interface modality, query interface, and result interface [2]. One traditional approach of presenting is to use popular Web browsers to navigate the IoT/WoT, but generic interfaces restrict interoperability between ecosystems. Visual Search for IoT and Map-based UI have also been proposed. Query UI, Results UI, and Dual-integrated interfaces must be distinguished. Structured Web forms or map-based interfaces are used for location-aware queries as input UI mechanisms. RESTful API is commonly used in WoTSE systems for M2M interactions. Other works suggest using specialised APIs, like SOAP, while RDF API is highlighted for manipulating semantic information in RDF graphs.

### C. Semantic Enrichment for IoT/WoT

SPITFIRE [50] offers a complete semantic enrichment for IoT/WoT scenarios through its search engine that utilises a vocabulary integrating data with Linked Open Data, an ontology describing entities and sensors, and a mechanism for semi-automatic sensor description. Mietz et al. [79] semantic model incorporates previous concepts for IoT search and includes a base sensor model, its states, history, prediction model, and placement. Various vocabulary models are available for enhanced search through relationships using RDF, RDFa, OWL, OWL-DL, or OWL-S in IoTSE/WoTSE systems.

Several strategies have been proposed for discovering semantically enabled smart things on the Web, including DiscoWoT, a search mechanism with load-balancing and query caching, the Web Avatar abstraction, and a user-system view around the Things Description model. Other approaches include a three-step search process with semantic profiles and an OWL-DL model for the IoT Ecosystem. Sense2Web provides an H2M interface for location-based search, while WoTSE is a search engine for the Semantic WoT.

All in all, semantic enrichment is used to address IoT interoperability challenges. The eWoT approach extends the W3C Thing Description model through semantic enrichment using RDF triple stores. This approach avoids the need for SPARQL support at IoT endpoints. Various ontologies have been used, including FOAF, SSN Ontology, and O&M Sensor Observation Ontology, to manage the vast amount of data produced by sensors. Dolce Ultralite, GeoName, and Phenonet Open IoT Ontology have also been effective in this domain.

## IV. TAXONOMIES OF SEARCH ENGINES FOR IoT—WoT BY DIMENSIONS

Taxonomies categorise IR systems for IoT and WoT using identification, characterisation, classification, and naming as key factors. Multiple dimensions are usually considered to propose the naming conventions.

From previous works, we highlight the taxonomy of Tran et al. [2] containing (24) dimensions, including *meta-path*. Tran et al. [2] added a fundamental analysis of

TABLE II  
TAXONOMY FRAMEWORKS, DIMENSIONS, AND DIRECTIONS

Taxonomy	Type	No. Studied Proposals	Assessed Dimmensions	Focus	Caveats
Multi-Taxonomy (Our Proposal)	IR Perspective	251	23	Holistic View	Diversity
Tran et al [59]	Flow-based for WoT	214	24	Modular Model	No IR in Depth
Pattar et al [81]	Data/Object-Centric	60	7	Applicability	No IR in Depth
Suparna et al [82]	Context-Model	55	6	Ontologies	No Characterisation
Noura et al [83]	Device-Platform-Grammar	31	12	Interoperability	No IR-oriented
Zhou et al [84]	Principles-Representation-Scope	49	9	Techniques	No Characterisation
Faheem et al [80]	Principles-Representation-Scope	37	4	Dynamic Search	No Characterisation
Aziez et al [60]	Architecture-Representation	23	10	Discovery	No IR-oriented
Romer et al [85]	Architecture Design	7	9	Real-time Local	Limited Revision
Zhang et al [86]	Architecture Design	6	14	Real-time Local	Limited Revision
Choudhury et al [87]	Architecture Design	5	6	Sensor Search	Limited Revision
Evdokimov et al [88]	Scope-Quality	5	9	Discovery	Limited Revision

approaches from a data/information flow viewpoint. One cornerstone piece of their taxonomy is the *meta-path*: classification in the form of a naming convention of IoTSE/WoTSE systems similar to a fingerprint with a strong focus on capturing the expected data/information flow. Nevertheless, it lacks an in-depth analysis from the IR perspective. The extensive dimension granularity biases the features rather than formulating a unified criterion and IR directions for evolution. In contrast, Rather than spreading the granularity of dimensions, our proposal aims to group them into IR stage subtasks. Faheem et al. [80] consolidated four dimensions: the use case (considered a dimension itself), thing schema, indexing, and ranking. Here, the substantial consolidation does not permit a precise characterization of the different works. The other taxonomies are in the middle, with dimensions covering the main IR subtask and experimental or prototype features.

Our proposal groups dimension into IR stage subtasks for precise characterization. The taxonomy aims to offer a holistic view of all IR approaches and insights on future evolution. In this section, we answer positively RQ1: Is/Are there any other SLR and taxonomies developed for IR-IoT and IR-WoT?, providing a list of existing taxonomies associated with IoTSE/WoTSE research, see Table II. We also present a perspective under the *Type* column, in which we group the taxonomy framework closely related to its focus and caveats. It contains the number of studied proposals reported (or the number of references) and the assessed dimensions.

We have found inconsistencies between classification models, mostly due to the different lenses adopted for the creation of these taxonomies while building the different perspectives (listed as follows).

- 1) Function and principles-based [84], [88].
- 2) Search scope and things model-based [81], [84].
- 3) Flow-based and architecture [59], [85], [86].
- 4) Application-specific and use case-based [80].

There is neither a common language nor a common understanding in classifying the IoTSE/WoTSE organisms and their constitutive parts. Furthermore, there is no IR-oriented perspective driving the classification. Even though previous works have evaluated dozens of parameters and dimensions, there is no holistic view that puts all the pieces together.

We compare all classifications in the primary studies dataset. We also present a comparison between existing classification models in Table II.

#### A. Proposed Unified and Holistic Multitaxonomy Framework

Our proposed framework (Table III) distinguishes between IoTSE and WoTSE systems. The former is intended for machine-to-machine interaction, while the latter is socially aware and offers a Web abstraction to return geo-location and perform predefined actions. Our taxonomy provides a precise overview of IR possibilities aimed at providing accurate information regardless of search techniques and models used. IoTSE and WoTSE are two distinct proposals for taxonomy. Although IoTSE has been extensively studied, WoTSE still requires sufficient identification and investigation. Recent works have not been classified or included in previous taxonomies due to the publishing time. All previous taxonomies universally include IoT data, stream, and content-based search. However, context awareness is not explicitly considered in the meta-path model in [2]. Another crucial aspect is the explicit recognition of IoT Predictive Search as part of the taxonomy framework. Although this could be understood as a technique to reduce the search scope, we believe that the potential of approaches in this category can shape the evolution of IR research.

Location-based IoT search is a distinct category that requires separate surveys and taxonomy due to its numerous subbranches. It is important to consider temporal context separately from predictive search. WoTSE families can facilitate thematic awareness and multiple search scopes. Thing-centred or social-centred WoT Search is the next generation of search engines that support end-users in finding multiple features and functionalities concerning things based on social relationships and providing sharing capabilities. Security, Privacy, and Trust shall be integrated into all the IR subtasks. WoT Actions Search involves virtual and physical actions triggered by intangible virtual actions in the real world. WoT Progressive Search is a promising research direction that can gradually approach spatial-temporal dimensions. Ultimately, the ideal WoT Everything Search is a species that can locate everything, including synthetic emotions and sensations.



TABLE III  
PROPOSED IOTSE/WOTSE TAXONOMY FRAMEWORK

Denotation		Our proposal	Tran et al.	Pattar et al.	Zhou et al.
WoT IR Systems	WoTSE	WoT Everything Search	◆	NC	▲
		WoT Progressive Search	◆	NC	◆
		WoT Actions Search	●	R → T → F	◆
		WoT Secure Search	◆	NC	◆
		WoT Thing Search	●	R → T → R + D	●
		WoT Social-based Search	◆	NC	●
		WoT Multi-modal Search	●	D + R → T → R	◆
IoT Search & Discovery Systems	IoTSE	IoT Semantic Search	◆	NC	▲
		IoT Temporal Search	▲	D + R → T → R	◆
		IoT Location Search	◆	NC	●
		IoT Predictive Search	▲	D → D	▲
		IoT Service Search	●	F → F	●
		IoT Resource Search	●	R → R	●
		IoT Context-based Search	◆	NC	●
		IoT Content-based Search	●	D (+ R) → T → R	●
		IoT Data Stream Search	●	D → D	●
		IoT Data Search	●	S → S	●

●	Fully Considered
▲	Partial Considered
◆	Not Considered

### B. Practical Applications and Real-World Scenarios

In this section, we describe several examples in the form of case studies and real-world scenarios which provide practical applications for the proposed taxonomy, helping validate the taxonomy's flexibility and effectiveness in real-world IoT/WoT environments, while guiding the applicability, development and research of new IR systems in IoT/WoT.

Some real-world scenarios are as follows.

#### 1) Smart City Traffic Management:

- a) *Description:* In a Smart City, real-time data from traffic sensors, smart vehicles, IoT devices, and User reports can be aggregated to manage traffic flow automatically.
- b) *Taxonomy Usage:* This application could employ multiple taxonomy elements.
  - i) IoT Temporal Search for analyzing traffic patterns and changes over time.
  - ii) IoT Location Search to retrieve data based on specific geographic areas (e.g., congestion in a particular intersection).
  - iii) IoT Predictive Search to forecast potential congestion or traffic incidents based on historical data and real-time updates.
- c) *Benefit:* This approach could help city managers, end-users (citizens), or automated systems proactively manage traffic, reduce congestion, and enhance urban mobility.

#### 2) Healthcare Monitoring and Assistance for Elderly Patients:

- a) *Description:* Healthcare providers can monitor elderly patients through wearable sensors, environmental IoT devices, and medical records to provide real-time assistance and support.
- b) *Taxonomy Usage:* This application could employ multiple taxonomy elements.

- i) IoT Context-based Search to filter alerts relevant to each patient's specific conditions and current context (e.g., heart rate anomaly only for patients with cardiac issues).
- ii) IoT Data Stream Search for continuously monitoring real-time vital signs data,
- iii) WoT Actions Search to trigger alerts or automated interventions if certain thresholds are breached (e.g., notifying caregivers) using Healt WoT-like abstractions.

- c) *Benefit:* This approach improves patient safety and responsiveness by ensuring quick access to and action on relevant patient data.

To illustrate the practical value and effectiveness of our proposed taxonomy, we present two case studies, showcasing its versatility and robustness. The first case study examines energy management in smart grid systems, highlighting the taxonomy's role in predictive and resource-based searches to optimize energy distribution and manage demand fluctuations across IoT-enabled devices and WoT-integrated appliances. The second case study focuses on personalized customer support within IoT-enabled smart homes, demonstrating how secure, proactive searches can enhance user satisfaction and system efficiency. All in all, these case studies validate the taxonomy's adaptability and provide a foundation for developing advanced IR systems tailored to the dynamic needs of IoT and WoT landscapes.

#### 1) Case Study (Real-Time Energy Management in Smart Grid Systems):

- a) *Objective:* Test how the taxonomy aids in managing and optimizing energy distribution based on demand patterns.
- b) *Scenario:* A smart grid system with IoT-connected meters and WoT-enabled appliances across residential and industrial areas is tasked with

- optimizing energy use and responding to demand fluctuations.
- c) *Taxonomy Demonstration:*
    - i) IoT Predictive Search to forecast future energy demand based on historical usage and current conditions.
    - ii) IoT Resource Search to locate and allocate available energy resources across the grid.
    - iii) WoT Multimodal Search to integrate data from different sources (e.g., weather forecasts, usage patterns).
  - d) *Outcome:* This study would measure how the taxonomy enables efficient search and retrieval, supporting demand-response strategies and improving grid resilience.
- 2) *Case Study (Personalized and Proactive Customer Support in IoT-Driven Smart Homes):*
    - a) *Objective:* Evaluate how the taxonomy supports delivering timely, contextually relevant information for proactive support in smart homes.
    - b) *Scenario:* In smart homes equipped with various IoT devices, a support system provides assistance based on device usage patterns and environmental context.
    - c) *Taxonomy Demonstration:*
      - i) IoT Semantic Search to understand and interpret the homeowner's actions and potential needs (e.g., a power surge in a kitchen appliance).
      - ii) WoT Actions Search to perform automated actions or suggest solutions.
      - iii) WoT Secure Search to ensure safe handling and retrieval of sensitive data for each user.
    - d) *Outcome:* This case study would test the taxonomy's ability to support intelligent, privacy-respecting, and proactive customer support, providing insights into enhancing user satisfaction and reducing costs.

## V. IOTSE—WOTSE EVALUATION FRAMEWORKS

Measuring the effectiveness of an IR system depends mostly on human evaluations of the usefulness of the information found and its relevance. IR, as a highly empirical discipline, requires careful and exhaustive evaluation to demonstrate the performance of its models [89]. There are several measures of interest that can be related to the quality of the responses in terms of the efficiency and effectiveness of IR for IoT—WoT. Traditionally: 1) precision and 2) completeness (*Recall*) are quality metrics used in IR and other related fields [81], [89]. From a more technical point of view, 3) the speed of response and 4) the size of the index are factors that can increase the quality of a user's experience with an IR system [33].

The importance of finding efficient and effective search methods for WoT applications has been identified by [84] as a commitment of all stages of IR. For example, Zhou et al. [84] has proposed multiple qualitative variables as metrics to evaluate search techniques in IoT—WoT applications: Query

Time, Query Precision. In terms of the response time of the IR system (which includes sending a query and receiving the results), as well as the precision of the results obtained. Thus, in the application of IR in IoT—WoT, evaluation methods that consider its multiple dimensions are also of vital importance.

### A. Classical IR Evaluation by Test Collection

To measure the effectiveness of IR in an ad-hoc way, traditionally, one has a collection of documents, a set of information needs tests, expressible as queries, and a set of relevance judgments, usually a binary evaluation of *relevant* or *not relevant* for each query-document pair. The availability of dynamic test collections focused on IoT—WoT paradigms [90] has been widely identified as a challenge and current need. None of the major IR evaluation forums, such as TREC,<sup>4</sup> NTCIR,<sup>5</sup> or FIRE<sup>6</sup> contains some specific collection for IoT—WoT paradigms.

In IR, it is possible to determine the effectiveness of a system on a set of topics using a test collection in conjunction with its respective specialized judgment of relevance [89]. From the results and the judgment, the following are determined: Precision (P) as the fraction of recovered documents that are relevant, *Recall* (R) as the fraction of relevant documents that are recovered and (F) as the harmonic mean weighted between precision and recall [81], [89]

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$F(1) = \frac{2 * TP}{2 * TP + FP + FN} \quad (3)$$

where TP represents true positives, FP represents false positives, FN represents false negatives, and TN represents true negatives. These can be adjusted to the type of recovery scheme, and there are variations, such as K-precision, R-precision, and mean average precision (MAP).

### B. Evaluation Frameworks, Datasets, and Test Collections for IoT—WoT

The SLR has been extended to cover a fifth group of questions related to the evaluation of IR systems in the IoT—WoT field.

- 1) How have the solutions proposed in the analyzed SLR studies evaluated the performance of IR-IoT and IR—WoT systems?
  - a) What time performance measures have been used?
  - b) What index size measurements have been used?
  - c) What classical evaluation measures, such as precision, completeness, and F-measure have been used?
  - d) Are there *datasets* or open data for reproducibility of experiments or reusability of data?

<sup>4</sup>Evaluation Tracks in TREC - <https://trec.nist.gov/data.html>.

<sup>5</sup>Evaluation Tracks in NTCIR - <http://research.nii.ac.jp/ntcir/data/data-en.html>.

<sup>6</sup>FIRE Evaluation Tracks - <http://fire.irs.ri.res.in/fire/static/data>.

TABLE IV  
EVALUATION FRAMEWORKS IN IR SYSTEMS FOR IoT-WoT

Complexity Evaluation Metrics			IR Evaluation Metrics			Evaluation Scenarios	
Time-related	Query-Related	Space	(P)	(R)	(F)	Objective	Related Work
Computational Time (ms)	No	Space Consumption	No	No	No	Scalability	Aguzzi et al. [11]
Execution Runtime (ms)	No	Index Size (B)	No	No	No	Generic Queries	Mili-Rodin [15]
System Speed (QPS/t)	Yes	No	No	No	No	Generic Queries	Hatcher et al. [14]
Query Volume (QV/t)	Yes	No	No	No	No	Specific Queries	Hatcher et al. [92]
Query Execution (t/Q)	Yes	No	No	No	No	Specific Queries	Iggena et al. [17]
Query Execution (t/Q)	Yes	No	No	No	No	Specific Queries	Cimmino et al. [93]
Query Processing Time (s)	Yes	No	Yes	Yes	No	Generic Queries	Pattar et al. [22]
Query Accuracy	Yes	Index Size (B)	Yes	Yes	No	Generic Queries	Tan et al. [7]

Table IV lists the main evaluation frameworks analyzed in the context of proposals for IR systems or search engines in IoT—WoT. Most *datasets* curated in the IoT sphere that have been found have a focus toward application to a specific field, such as testing security mechanisms. Such is the case of the *dataset* BoT-IoT.<sup>7</sup> This contains around 72 million records with the respective categorization and/or automatic relevance of attack type and other technical details. Another source of information consulted is the Google data search system, available at <https://datasetsearch.research.google.com/>. Where you can find a few *datasets* created by recent research in the WoT and IR field independently.

In the IoTSE evaluation presented by [91], IoTSEs underwent a query of IoT-enabled sensors that measure apparent temperature in degrees Celsius. All instances achieved perfect accuracy and completeness due to the use of a database to store and resolve queries on the sensor metadata. The design of efficiency evaluation methods in the implementation and use of IoT search is a topic of vital importance [92]. Hatcher et al. propose that the efficiency of the search system in IoT be developed from two perspectives, increasing the effectiveness in search of each component in processing time and the scalability in the interactions between RI stages by increasing the “Throughput” or capacity of processing given in queries per second.

A significant majority of IoTSE—WoTSE proposals base the evaluation of performance on real complexity in time and space. The time measures are associated with the processing of the index, queries, or the entire recovery stages until the delivery of results to the end user. And the space measurements associated with the size of the index or the storage used.

Recently, Cimmino and García-Castro [93] reported the results of some experiments carried out for the semantic discovery of WoT, framed in the WoT W3C initiative. This uses the formal TD definitions and makes use of ten SPARQL type queries to analyze the query time based on the number of things (1–1000 TDs).<sup>8</sup> Another open *dataset* related to the W3C initiative is the one released by Chapernay and Käbisch [94]. This work is based on the W3C TD thing description model and proposes the construction

of an ontology (W3C TD Ontology)<sup>9</sup> for modeling the real world from a collection of virtual things. The released dataset comprises less than a hundred things modeled in JSON, with sixty-five things in total.

### C. Creating IR-IoT—IR-WoT Test Collection

Given the evaluation needs of IR systems for IoT—WoT paradigms and the lack of open *datasets* as well as test collections focused on IoT—WoT paradigms, the following paths are available in order to provide experiments reproducibility mechanisms and data reusability in future work in the study, analysis, and development of IR systems in the field. And ultimately build an evaluation framework for RI-IoT and RI-WoT systems.

- 1) Creating an IR-IoT—IR-WoT test collection.
  - a) Starting with the modification of Collections for the evaluation of IR performance with a real-time approach.
  - b) From IoT Dataset oriented to time series.
  - c) From WoT Dataset oriented to semantics.
- 2) Construction via active learning of an IR test collection [95].
- 3) IR evaluation as a search simulation, one of the proposals at the recent NTCIR 2022<sup>10</sup> and at [96].

### D. Evaluation and Relevance Judgments

Creating a classical test IR collection requires great effort in obtaining user evaluations and relevance judgments by experts. Methodologically speaking, relevance assessments should be compiled considering the following.

- 1) Information Needs expressed in the form of Queries. Which determines a number of queries  $Q$ , on a certain number of topics  $T$ .
- 2) A set of documents  $(N, K)$  retrieved by different IR systems  $(A—B—C)$ .
- 3) A relevance evaluation is given by Experts on the topic(s) of the recovered documents (relevance judgments).

During the evolution of IR research, different measures have been proposed to characterize the agreement between judges and their evaluation of the relevance of the documents in the collection, using the so-called Kappa statistics [89].

<sup>7</sup>Dataset BoT-IoT - <https://research.unsw.edu.au/projects/toniot-datasets>.

<sup>8</sup>Dataset and Open Experiment Data - <https://doi.org/10.5281/zenodo.6674151>

<sup>9</sup>Draft Thing Description Ontology - <https://www.w3.org/2019/wot/td>

<sup>10</sup>Keynote at NTCIR 2022 - <http://research.nii.ac.jp/ntcir>.

## VI. OPEN RESEARCH DISCUSSIONS AND FUTURE WORK

Multiple challenges have been documented in the literature during the last years; some persist, while others, through time, have been diminished by novel solutions (Table V). Existing state-of-the-art surveys have well-documented challenges and future needs of IoTSE/WoTSE and their inherent retrieval capabilities, focusing on a specific subject or the impact of the challenges in some application fields or technical factors. This section presents a review of the challenges from architectural and technical perspectives.

### A. Architectural Challenges

1) *Dynamicity and Dynamics*: The biggest architectural challenge is the high pace of IoT/WoT data generation and changes in things' states and sensors' data. Recently, Faheem et al. [80] presented dynamic searching as the main current challenge faced by IoTSE/WoTSE. Due to that, most WoTSE approaches are based only on keyword-based search or looking for static locations. In that way, it urges the construction of dynamic indexing mechanisms and consideration of intent-based ranking techniques, which can positively impact the dynamicity, adaptability, and scalability of WoTSE. Meriem et al. [60] defended the idea that dynamicity is core to finding the relevant service meeting the end-user requirement in real-time. While Ma and Liu [27] expressed the dynamicity of search as guaranteed timely results in terms of freshness. All in all, [2] in agreement with [80] points out the open issues of dynamic searching resulting in a tradeoff between indexing and freshness.

2) *Adaptability of IR*: It refers to the ability of a WoTSE to adapt its behavior to different IoT/WoT scenarios, including the modification of internal IR components to fulfill IoT/WoT demands. Lately, Meriem et al. [60] and Aziez et al. [97] addressed the IoT discovery problem by providing a classification of services based on technical features and relationships, which is extensive, depicting a complete comparative study. It highlights adaptability as an existing challenge because most approaches do not address the widespread IoT requirements that need to be met by classical IR techniques. Tran et al. [2] agreed on the discussion reporting that every stage of the IR process for an ideal WoTSE shall require some kind of adaptation. One of the biggest steps would be to isolate the heterogeneity of data and IoT technologies by middleware. IR processing parallelism is identified as a useful strategy when dealing with data analysis and IR, considering the amount of generated data, if it is voluminous and independent [98].

3) *Conceptual Heterogeneity*: Four main obstacles shadow the evolution of the field due to the heterogeneity of concepts.

*Difficult-to-Reproduce* due to the nonexistence of IoT/WoT extended datasets [59], which in the majority of works are proprietary or private. It is required to build open and public datasets for IoTSE/WoTSE research. Some public and well-known datasets are indeed being used; nevertheless, they are more oriented toward sensing scenarios rather than providing the whole picture. Moreover, full experiments can not be replicated, whether due to a lack of datasets or problems with

the reproducibility of the experiments. The reproducibility problem is related to three main factors.

- 1) *Lack of Dataset*: Use of proprietary dataset or use of subsample of the publicly available dataset that has not been published or for which the subsampling has not been disclosed.
- 2) Lack of details about the evaluation pipeline (processing of data) and protocol (tasks).
- 3) Lack of details about the system parameters.

*Difficult-to-Evaluate/Compare* is an implication of the previous point, which shadows the evolution of IR systems, given the divergence in the evaluation criteria of performance metrics. A significant set of works takes as essential time-based metrics for evaluating their approaches. However, we should rely on adjusted or modified versions of IR metrics for comparing the performance not only in terms of efficiency (time and space complexity) but also in terms of effectiveness, as offline and online IR metrics. As per our literature revision, there are no documented efforts to standardize the evaluation process by creating some evaluation protocols or adopting some existing approaches.

*Difficult-to-Access* is another important point frequently mentioned. We witness optimistically new collections of datasets published in Public Clouds, such as Google Cloud Platform<sup>11</sup> with nearly 200 datasets. Some examples of live and dynamic datasets are the Chicago Taxi Trips and the NYC TLC Trips. Also, Amazon AWS provides open access to 175 datasets<sup>12</sup> mainly oriented to medical and spatial purposes. However, the availability of datasets is not the only blocking point, as there are still many domains and data that are not open to researchers.

*Difficult-to-Reuse*: Tran et al. [59] noticed different implementations for similar modular components: the difficulty of reuse components comes from the use of not typical architectures and interfaces between components. Reusability can be achieved through the construction of agreed architectures, descriptors, and libraries. Standardization becomes a structural pillar of the next generation IoTSE/WoTSE.

### B. Technical Challenges

1) *Scalability*: It refers to adjusting the IoTSE/WoTSE computing resources to handle the colossal amount of things and produced data. Pattar et al. [81] and Zhang et al. [86] identified the architectural design of IoTSE as the core component to be adapted with new solutions. Technically speaking, scalability is seen as the ability of an IoT-related system to adapt to changes in the real world environment and meet future needs. An IoTSE/WoTSE must be capable of handling the growing workload in terms of processing, storage, and communications capacities.

2) *Interoperability*: refers to the ability of IoTSE/WoTSE to interchange information with other systems, to be modular with no dependencies between layers, facilitating innovation and evolution. Furthermore, to yield standards IoT/WoT

<sup>11</sup><https://cloud.google.com/public-datasets>

<sup>12</sup><https://registry.opendata.aws/>

TABLE V  
SHORT AND LONG-TERM RESEARCH TOPICS FOR IR-IoT AND IR-WoT

Research Topic	Term	Research Description
Experimental Frameworks	Short	Experimental frameworks have been presented through the literature. Special focus is required in convergence or in providing accessibility/re-usability, reproducibility of the evaluation criteria among all the approaches. Those are the pillar fundamentals required for powering and stimulating IR-IoT/WoT development and evolution.
Common Evaluation Criteria	Short	We agree with Tran et al. [59], [90] perspective that IR research for IoT and WoT is a complex area with technical and standardisation challenges to overcome. Comparability of works requires not only standardisation of concepts and models, but more importantly a common evaluation framework similar to the TREC initiative <a href="https://trec.nist.gov/trec.tracks.html">https://trec.nist.gov/trec.tracks.html</a> .
IR in the Social WoT	Long	Khadir et al. [102] denoted as social avatars, a particular type of WoT avatars able to be linked to each other according to commitments and similarity criteria, providing collective intelligence for more precise answers and WoT service discovery. IR mechanism dealing with the social context type of queries respect to the social avatars will be required.
Progressive and Personalised Search based on Emotions and Senses	Long	It is argued that semantic enrichment focuses on finding similarities but not on providing personalised user-centric results [22]. Progressive mechanism shall aim to reduce the search space while reducing the query time systematically. At the same time some semantic enrichment should personalise user results. Going beyond the traditional boundaries of information retrieval, Luis-Ferreira et al. [103] propose a multi-modal search including sensations and emotions as conductors for searching things, mimicking the way our brains store and find things by experience. It provides a clear and valid contribution to the sensorial description of things and the multi-modal futuristic search required in future research.

models [83], points to interoperability as an umbrella of issues that slows down the evolution and prevents IoT emergence. Interoperability is present at different levels: network, device, syntactic, semantic, and platform. Some open challenges permeate the creation of cross-domain and cross-platform composited applications. It is mandatory to mature the protocol standardization for device-to-device communications, the openness of API approaches, and the unification of testing and evaluation frameworks. In the end, a lack of interoperability in the IoT/WoT layer will undoubtedly impact the IoTSE/WoTSE evolution as well. Interoperability remains an open issue for IoT/WoT and Semantic WoT as well [24]. Even though semantics promises to solve the interoperability issues of multiple approaches, protocols, integration, and applications, it involves defining the role of each entity and element in the IoT/WoT. With no consensus about the ontologies and knowledge representation, the interoperability advantage is blurred and fuzzy. Silva et al. [24] presented a different approach using a different language, SWoTPADL, rather than OWL-S, extended from WSLM. SWoTPADL aims to provide a service composition engine for SWoT apps and service mashups.

3) *Security, Privacy, and Trust*: These are seen as critical challenges as IoTSE/WoTSE data shall be protected, confidential, and private in all IR tasks. Security, Privacy, and Trust have been relatively unexplored issues. Since 2020, we recognize major efforts to cover security, privacy, and trust on IoTSE/WoTSE. Yang et al. [21] proposed a Participant Selection Strategy With Privacy Protection (PSSPP) for IoTSE. It provides anonymity and mixed mechanisms for end-users and requests at query time. The PSSPP system evaluates participants' trust value and credibility in the IoT search at mobile crowdsensing scenarios. A minority of works are increasing the research attention on adhering to the security, privacy, and trust dimensions of IoTSE/WoTSE. Barclay et al. [99] proposed discoverable trusted services

in highly dynamic workflows, like those used in 5G/WoT-like scenarios. It provides an enhanced semantic search space for efficient and trusted 5G/WoT-like service discovery. Other recent works are specific to some IoT/WoT industrial/health applications. Liu et al. [100] proposed a multikeyword searchable encryption scheme for electronic health files (EHFs) in the Medical IoT (MIoT). It provides fine-grained access authorization and sharing mechanisms for EHFs in MIoT. In the same context of MIoT, Bao et al. [101] proposed a lightweight attribute-based searchable encryption scheme with fine-grained access control and authorization, allowing a keyword-based search. More recently, Hatcher et al. [14] presented a study of security issues, challenges, and vulnerabilities in IoTSE systems in conjunction with a taxonomy of those. Recently, I-Recon [10], a new search engine, addressed the limitations of existing public IoTSE-like tools like Shodan and Censys by offering advanced search capabilities, customizable scanning parameters, real-time incident response, and detailed error tracking. It enables complex queries, aggregation searches, and efficient metadata filtering, enhancing vulnerability analysis and threat detection as a unique characteristic over the whole spectrum of IoTSE systems.

### C. Evolutionary Directions and Engineering

The modular architecture presented by [59] and the proposed framework for component-based IoTSE in [90] constitute two consistent starting points for the design, engineering and construction of evolutionary IoTSE/WoTSE systems. Moreover, it is noticed that similar internal IR functional blocks are being adapted to different types of IoT/WoT content. We argue that design decisions are being affected by the Thing description and the semantic mechanisms. From the architectural point of view, a challenge is to build a generalised WoTSE able to perform both local and global search. This can be achieved with a distributed approach taken to the edge

and then interlinked through federation, integrating IoTSE with edge network and computing techniques, co-designing the evolution of Cloud and IoT. Through standardisation, IoTSE/WoTSE shall be able to be extended to manage multiple interactions with cyber-physical systems and not only be application-specific to a unique scenario. There is room for performance optimisation given the impact of IoT and WoT on the heterogeneity, dynamicity, and scalability dimensions. These remain significant challenges for the full development and evolution of IoT/WoT systems [28]. Finally, but not least important, security, privacy and trust should be fully integrated and addressed by IoTSE/WoTSE.

## VII. CONCLUSION

As the Internet and WoT become more pervasive, new search mechanisms are needed to handle the real-time data collected by IoT devices. However, there is a lack of agreement on how to model, name, and characterize these devices, hindering interoperability. Semantic enrichment through ontologies can help, but challenges remain due to multiple perspectives.

We analyzed the challenges of IoTSE and WoTSE and identified remaining issues in IR models for IoT and WoT. New approaches are emerging to solve problems in discovering, crawling, indexing, and ranking information. Our contribution is a proposed taxonomy and survey of the state-of-the-art in IR systems for IoT and WoT. Furthermore, we have zoomed in on every IR stage with a separate discussion and subclassification. It allows apprehending how the adaptation of IR techniques and methodologies to specific application scenarios can be a way to overcome the existing challenges and open issues.

A typical IoTSE+WoTSE architecture can contribute to the development of new evolutionary directions for the following generation of Search Engines. Our ultimate goal is to bring to light the different perspectives and correlate them in a holistic, 360° panoramic of approaches that have been or are being developed, homogenizing terms and shared understanding.

We firmly believe that dynamicity, adaptability, heterogeneity, and scalability directly impact IR and Search Engine systems' conception, as pointed out by the vast body of research for IoT-IR/WoTSE-IR. The analysis of these studies has pointed out a lack of standards and common directions for IoTSE/WoTSE. Therefore, it is essential to fulfill the requirements for a common taxonomy that will help to identify the evolution of the IR field, with its main approaches and relationships, and, in the future, to guide through comparative analysis.

## ACKNOWLEDGMENT

Funding for open access charge: Universidad de Granada / CBUA.

## REFERENCES

- [1] M. Mrissa, L. Médini, J. Jamont, N. Le Sommer, and J. Laplace, "An avatar architecture for the Web of Things," *IEEE Internet Comput.*, vol. 19, no. 2, pp. 30–38, Mar./Apr. 2015.
- [2] N. K. Tran, Q. Z. Sheng, M. A. Babar, and L. Yao, "Searching the Web of Things: State-of-the-art, challenges, and solutions," *ACM Comput. Surveys*, vol. 50, no. 4, pp. 1–34, Aug. 2017. [Online]. Available: <https://doi.org/10.1145/3092695>

- [3] W. E. Zhang, Q. Z. Sheng, A. Mahmood, D. H. Tran, M. Zaib, and S. A. Hamad, "The 10 research topics in the Internet of Things," in *Proc. IEEE 6th Int. Conf. Collaboration Internet Comput. (CIC)*, vol. 1, New York, NY, USA, 2020, pp. 34–43.
- [4] M. A. Mohamed, G. Kardas, and M. Challenger, "Model-driven engineering tools and languages for cyber-physical systems—A systematic literature review," *IEEE Access*, vol. 9, pp. 48605–48630, 2021.
- [5] B. Ostermaier, K. Römer, F. Mattern, M. Fahrmaier, and W. Kellerer, "A real-time search engine for the Web of Things," in *Proc. Internet Things (IOT)*, 2010, pp. 1–8.
- [6] H. Wang, C. C. Tan, and Q. Li, "Snoogle: A search engine for pervasive environments," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 8, pp. 1188–1202, Aug. 2010.
- [7] C. C. Tan, B. Sheng, H. Wang, and Q. Li, "MicroSearch: A search engine for embedded devices used in pervasive computing," *ACM Trans. Embedded Comput. Syst.*, vol. 9, no. 4, pp. 1–29, Apr. 2010. [Online]. Available: <https://doi.org/10.1145/1721695.1721709>
- [8] H. Cheng et al., "MSNet: Structural wired neural architecture search for Internet of Things," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, 2019, pp. 2033–2036.
- [9] R. Krishankumar, F. Ecer, A. R. Mishra, K. S. Ravichandran, A. H. Gandomi, and S. Kar, "A SWOT-based framework for personalizing ranking of IoT service providers with generalized fuzzy data for sustainable transport in urban regions," *IEEE Trans. Eng. Manag.*, vol. 71, pp. 2937–2950, 2024.
- [10] V. Moghissi and A. Shamel-Sendi, "I-RECON: An IoT-based search engine for Internet-facing services vulnerability reconnaissance," *IEEE Access*, vol. 12, pp. 96100–96112, 2024.
- [11] C. Aguzzi, L. Gigli, I. Zyrianoff, and L. Roffia, "ZION: A scalable W3C Web of Things directory," in *Proc. IEEE 21st Consum. Commun. Netw. Conf. (CCNC)*, 2024, pp. 1–6.
- [12] H. Chen, L. Wang, W. Qin, X. Zhou, and L. Cui, "VSA-SD: A service discovery method based on vector symbol architecture for low-cost IoT system development," *IEEE Trans. Cloud Comput.*, vol. 12, no. 1, pp. 145–158, Jan.–Mar. 2024.
- [13] M. R. Faheem, T. Anees, M. Hussain, A. Ditta, H. Alquhayz, and M. A. Khan, "Indexing in WoT to locate indoor things," *IEEE Access*, vol. 11, pp. 53497–53517, 2023.
- [14] W. G. Hatcher, C. Qian, F. Liang, W. Liao, E. Blasch, and W. Yu, "Secure IoT search engine: Survey, challenges issues, case study, and future research direction," *IEEE Internet Things J.*, vol. 9, no. 18, pp. 16807–16823, Sep. 2022.
- [15] M. S. Eddine and V. Rodin, "WoT search engine based on multi agent system: A conceptual framework," *Int. J. Interactive Mobile Technol.*, vol. 16, no. 5, pp. 49–61, Mar. 2022. [Online]. Available: <https://online-journals.org/index.php/i-jim/article/view/27901>
- [16] N. Pavlopoulou and E. Curry, "PoSSUM: An entity-centric publish/subscribe system for diverse summarization in Internet of Things," *ACM Trans. Internet Technol.*, vol. 22, no. 3, p. 73, Mar. 2022. [Online]. Available: <https://doi.org/10.1145/3507911>
- [17] T. Iggena et al., "IoT-Crawler: Challenges and solutions for searching the Internet of Things," *Sensors*, vol. 21, no. 5, pp. 1–32, 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/5/1559>
- [18] R. R. Nandan, N. Nalini, and P. N. Hamsavath, "IoT-CBSE: A search engine for semantic Internet of Things," in *Emerging Research in Computing, Information, Communication and Applications*, N. R. Shetty, L. M. Patnaik, H. C. Nagaraj, P. N. Hamsavath, and N. Nalini, Eds. Singapore: Springer, 2022, pp. 265–271.
- [19] L. Sciuillo, F. Montori, A. Trotta, M. Di Felice, and T. S. Cinotti, "Discovering Web Things as services within the arrowhead framework," in *Proc. IEEE Conf. Ind. Cyberphys. Syst. (ICPS)*, vol. 1, 2020, pp. 571–576.
- [20] J. Tang, Z. Zhou, X. Xue, and G. Wang, "Using collaborative edge-cloud cache for search in Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 2, pp. 922–936, Feb. 2020.
- [21] P. Yang, X. Kang, Q. Wu, B. Yang, and P. Zhang, "Participant selection strategy with privacy protection for Internet of Things search," *IEEE Access*, vol. 8, pp. 40966–40976, 2020.
- [22] S. Pattar et al., "Progressive search algorithm for service discovery in an IoT ecosystem," in *Proc. Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber Phys. Soc. Comput. (CPSCoM) IEEE Smart Data (SmartData)*, 2019, pp. 1041–1048.
- [23] L. Sciuillo, C. Aguzzi, M. Di Felice, and T. S. Cinotti, "WoT store: Enabling things and applications discovery for the W3C Web of Things," in *Proc. 16th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, 2019, pp. 1–8.

- [24] A. L. M. Silva, J. d. J. Pérez-Alcázar, and S. T. Kofuji, "Interoperability in semantic Web of Things: Design issues and solutions," *Int. J. Commun. Syst.*, vol. 32, no. 6, 2019, Art. no. e3911. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/dac.3911>
- [25] J. Tang, Z. Zhou, L. Shu, and G. Hancke, "SMPKR: Search engine for Internet of Things," *IEEE Access*, vol. 7, pp. 163615–163625, 2019.
- [26] F. Karim, I. Lytra, C. Mader, S. Auer, and M.-E. Vidal, "DESERT: A continuous SPARQL query engine for on-demand query answering," *Int. J. Semantic Comput.*, vol. 12, no. 3, pp. 373–397, 2018. [Online]. Available: <https://doi.org/10.1142/S1793351X18400172>
- [27] H. Ma and W. Liu, "A progressive search paradigm for the Internet of Things," *IEEE MultiMedia*, vol. 25, no. 1, pp. 76–86, Jan.–Mar. 2018.
- [28] A. F. Skarmeta et al., "IoT-Crawler: Browsing the Internet of Things," in *Proc. Global Internet Things Summit (GloTS)*, 2018, pp. 1–6.
- [29] A. Shemshadi, Q. Z. Sheng, Y. Qin, A. Sun, W. E. Zhang, and L. Yao, "Searching for the Internet of Things: Where it is and what it looks like," *Pers. Ubiquitous Comput.*, vol. 21, no. 6, pp. 1097–1112, Dec. 2017. [Online]. Available: <https://doi.org/10.1007/s00779-017-1034-0>
- [30] L. H. Nunes et al. "A distributed sensor data search platform for Internet of Things environments." 2016. [Online]. Available: <http://arxiv.org/abs/1606.07932>
- [31] Y. Qin, A. Shemshadi, Q. Sheng, and A. Alzubaidi, "CEIoT: A framework for interlinking smart things in the Internet of Things," in *Proc. 12th Int. Conf. Adv. Data Min. Appl. (ADMA)*, Jan. 2016, pp. 203–218.
- [32] A. Shemshadi, Q. Z. Sheng, and Y. Qin, "ThingSeek: A crawler and search engine for the Internet of Things," in *Proc. 39th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval (SIGIR)*, 2016, pp. 1149–1152. [Online]. Available: <https://doi.org/10.1145/2911451.2911471>
- [33] M. Younan, S. Khattab, and R. Bahgat, "WoTSF: A framework for searching in the Web of Things," in *Proc. 10th Int. Conf. Inf. Syst. (INFOS)*, 2016, pp. 278–285. [Online]. Available: <https://doi.org/10.1145/2908446.2908496>
- [34] M. Ebrahimi, E. Shafteibavani, R. K. Wong, and C. Chi, "A new meta-heuristic approach for efficient search in the Internet of Things," in *Proc. IEEE Int. Conf. Services Comput.*, 2015, pp. 264–270.
- [35] P. Gomes, E. Cavalcante, T. Rodrigues, T. Batista, F. C. Delicato, and P. F. Pires, "A federated discovery service for the Internet of Things," in *Proc. 2nd Workshop Middleware Context-Aware Appl. IoT (M4IoT)*, 2015, pp. 25–30. [Online]. Available: <https://doi.org/10.1145/2836127.2836129>
- [36] M. Li, H. Chen, X. Huang, and L. Cui, "EasiCrawl: A sleep-aware schedule method for crawling IoT sensors," in *Proc. IEEE 21st Int. Conf. Parallel Distrib. Syst. (ICPADS)*, 2015, pp. 148–155.
- [37] J. Soldatos et al., "OpenIoT: Open source Internet-of-Things in the cloud," in *Interoperability Open-Source Solutions Internet of Things*, I. Podnar Žarko, K. Pripuzić, and M. Serrano, Eds. Split, Croatia: Springer Int., 2015, pp. 13–25.
- [38] R. Kolcun and J. A. McCann, "DRAGON: Data discovery and collection architecture for distributed IoT," in *Proc. Int. Conf. Internet Things (IOT)*, 2014, pp. 91–96.
- [39] J. Michel, C. Julien, and J. Payton, "Gander: Mobile, pervasive search of the here and now in the here and now," *IEEE Internet Things J.*, vol. 1, no. 5, pp. 483–496, Oct. 2014.
- [40] C. Perera, A. Zaslavsky, P. Christen, M. Compton, and D. Georgakopoulos, "Context-aware sensor search, selection and ranking model for Internet of Things middleware," in *Proc. IEEE 14th Int. Conf. Mobile Data Manag.*, vol. 1. New York, NY, USA, 2013, pp. 314–322.
- [41] A. J. Jara, P. Lopez, D. Fernandez, J. F. Castillo, M. A. Zamora, and A. F. Skarmeta, "Mobile discovery: A global service discovery for the Internet of Things," in *Proc. IEEE 27th Int. Conf. Adv. Inf. Netw. Appl. Workshops*, 2013, pp. 1325–1330.
- [42] Z. Ding, J. Dai, X. Gao, and Q. Yang, "A hybrid search engine framework for the Internet of Things," in *Proc. IEEE 9th Web Inf. Syst. Appl. Conf.*, 2012, pp. 57–60.
- [43] A. J. Jara, P. Martinez-Julia, and A. Skarmeta, "Light-weight multicast DNS and DNS-SD (ImDNS-SD): IPv6-based resource and service discovery for the Web of Things," in *Proc. 6th Int. Conf. Innov. Mobile Internet Services Ubiquitous Comput.*, 2012, pp. 731–738.
- [44] D. Le-Phuoc, H. Q. Nguyen-Mau, J. X. Parreira, and M. Hauswirth, "A middleware framework for scalable management of linked streams," *J. Web Semantics*, vol. 16, pp. 42–51, Nov. 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1570826812000728>
- [45] A. Pintus, D. Carboni, and A. Piras, "ParaiMPU: A platform for a social Web of Things," in *Proc. 21st Int. Conf. World Wide Web (WWW)*, 2012, pp. 401–404. [Online]. Available: <https://doi.org/10.1145/2187980.2188059>
- [46] X. Qian and X. Che, "Security-enhanced search engine design in Internet of Things," *J. Univ. Comput. Sci.*, vol. 18, no. 9, pp. 1218–1235, May 2012. [Online]. Available: [http://www.jucs.org/jucs\\_18\\_9/security\\_enhanced\\_search\\_engine](http://www.jucs.org/jucs_18_9/security_enhanced_search_engine)
- [47] M. Komatsuzaki, K. Tsukada, I. Siio, P. Verronen, M. Luimula, and S. Pieskä, "ITEMinder: Finding items in a room using passive RFID tags and an autonomous robot (poster)," in *Proc. 13th Int. Conf. Ubiquitous Comput. (UbiComp)*, 2011, pp. 599–600. [Online]. Available: <https://doi.org/10.1145/2030112.2030232>
- [48] D. Le-Phuoc, M. Dao-Tran, J. X. Parreira, and M. Hauswirth, "A native and adaptive approach for unified processing of linked streams and linked data," in *Proc. Semantic Web (ISWC)*, 2011, pp. 370–388.
- [49] S. Mayer and D. Guinard, "An extensible discovery service for smart things," in *Proc. 2nd Int. Workshop Web Things (WoT)*, 2011, p. 12. [Online]. Available: <https://doi.org/10.1145/1993966.1993976>
- [50] D. Pfisterer, K. Romer, D. Bimschas, O. Kleine, R. Mietz, and C. Truong, "SPITFIRE: Toward a semantic Web of Things," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 40–48, Nov. 2011.
- [51] B. M. Elahi, K. Romer, B. Ostermaier, M. Fahrmaier, and W. Kellerer, "Sensor ranking: A primitive for efficient content-based sensor search," in *Proc. Int. Conf. Inf. Process. Sensor Netw.*, 2009, pp. 217–228.
- [52] C. A. Henson, J. K. Pschorr, A. P. Sheth, and K. Thirunaryan, "SemSOS: Semantic sensor observation service," in *Proc. Int. Symp. Collaborative Technol. Syst.*, 2009, pp. 44–53.
- [53] S. Jirka, A. Bröring, and C. Stasch, "Discovery mechanisms for the sensor Web," *Sensors*, vol. 9, no. 4, pp. 2661–2681, Apr. 2009, doi: [10.3390/s90402661](https://doi.org/10.3390/s90402661).
- [54] C. Frank, P. Bolliger, F. Mattern, and W. Kellerer, "The sensor Internet at work: Locating everyday items using mobile phones," *Pervasive Mobile Comput.*, vol. 4, no. 3, pp. 421–447, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1574119207000764>
- [55] W. I. Grosky, A. Kansal, S. Nath, J. Liu, and F. Zhao, "SenseWeb: An infrastructure for shared sensing," *IEEE MultiMedia*, vol. 14, no. 4, pp. 8–13, Oct.–Dec. 2007. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/senseweb-an-infrastructure-for-shared-sensing/>
- [56] S. Nath, J. Liu, and F. Zhao, "SensorMap for wide-area sensor Webs," *IEEE Comput.*, vol. 40, no. 7, pp. 90–93, Jan. 2007. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/sensormap-for-wide-area-sensor-webs/>
- [57] K.-K. Yap, V. Srinivasan, and M. Motani, "MAX: Human-centric search of the physical world," in *Proc. 3rd Int. Conf. Embedded Netw. Sensor Syst. (SenSys)*, 2005, pp. 166–179. [Online]. Available: <https://doi.org/10.1145/1098918.1098937>
- [58] X. Li, Y. J. Kim, R. Govindan, and W. Hong, "Multi-dimensional range queries in sensor networks," in *Proc. 1st Int. Conf. Embedded Networked Sensor Syst. (SenSys)*, 2003, pp. 63–75. [Online]. Available: <https://doi.org/10.1145/958491.958500>
- [59] N. K. Tran, Q. Z. Sheng, M. A. Babar, L. Yao, W. E. Zhang, and S. Dustdar, "Internet of Things search engine: Concepts, classification, and open issues." 2018. [Online]. Available: <http://arxiv.org/abs/1812.02930>
- [60] A. Meriem, S. Benharzallah, and H. Bennoui, "A full comparison study of service discovery approaches for Internet of Things," *Int. J. Pervasive Comput. Commun.*, vol. 15, no. 1, pp. 30–56, 2019. [Online]. Available: <https://doi.org/10.1108/IJPC-04-2019-0038>
- [61] M. A. Ferrag, Z. Kouahla, H. Seridi, and M. Kurulay, "Big IoT data indexing: Architecture, techniques and open research challenges," in *Proc. Int. Conf. Netw. Adv. Syst. (ICNAS)*, 2019, pp. 1–6.
- [62] M. Liu, D. Li, Q. Chen, J. Zhou, K. Meng, and S. Zhang, "Sensor information retrieval from Internet of Things: Representation and indexing," *IEEE Access*, vol. 6, pp. 36509–36521, 2018.
- [63] J. Shin, S. Eom, and K. Lee, "Q-ASSF: Query-adaptive semantic stream filtering," in *Proc. IEEE 9th Int. Conf. Semantic Comput. (IEEE ICSC)*, 2015, pp. 101–108.
- [64] Q.-T. Doan, A. S. M. Kayes, W. Rahayu, and K. Nguyen, "A framework for IoT streaming data indexing and query optimisation," *IEEE Sensors J.*, vol. 22, no. 14, pp. 14436–14447, July 2022.
- [65] M. Younan, M. Elhoseny, A. E. M. A. Ali, and E. H. Houssein, "Data reduction model for balancing indexing and securing resources in the Internet-of-Things applications," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5953–5972, Apr. 2021.

- [66] S. Tang et al., "Coordinate-based efficient indexing mechanism for intelligent IoT systems in heterogeneous edge computing," *J. Parallel Distrib. Comput.*, vol. 166, pp. 45–56, Aug. 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0743731522000892>
- [67] I. Kemouquette, Z. Kouahla, A.-E. Benrazek, B. Farou, and H. Seridi, "Cost-effective space partitioning approach for IoT data indexing and retrieval," in *Proc. Int. Conf. Netw. Adv. Syst. (ICNAS)*, 2021, pp. 1–6.
- [68] C. Du, Z. Zhou, S. Ying, J. Niu, and Q. Wang, "An efficient indexing and query mechanism for ubiquitous IoT services," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 18, no. 4, pp. 245–255, Apr. 2015. [Online]. Available: <https://doi.org/10.1504/IJAHUC.2015.069060>
- [69] J. Tang and Z. Zhou, "Searching the Internet of Things using coding enabled index technology," in *Green, Pervasive, Cloud Computing*, S. Li, Ed. Hangzhou, China: Springer Int., 2019, pp. 79–91.
- [70] J. Tang, X. Xue, S. Yangui, and Z. Zhou, "Efficient search for moving object devices in Internet of Things networks," in *Proc. IEEE Int. Conf. Web Services (ICWS)*, 2020, pp. 454–462.
- [71] C. Diamantini, A. Nocera, D. Potena, E. Storti, and D. Ursino, "Querying the IoT using multiresolution contexts," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 6127–6139, Apr. 2021.
- [72] P. Zhang, X. Li, Y. Liu, X. Kang, and Y. Liu, "SDU: State-based dual-mode sensor search mechanism toward Internet of Things," *IEEE Access*, vol. 7, pp. 147962–147974, 2019.
- [73] M. Liu, D. Li, Y. Zeng, W. Huang, K. Meng, and H. Chen, "Combinatorial-oriented feedback for sensor data search in Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 284–297, Jan. 2020.
- [74] C. Perera, A. Zaslavsky, C. H. Liu, M. Compton, P. Christen, and D. Georgakopoulos, "Sensor search techniques for sensing as a service architecture for the Internet of Things," *IEEE Sensors J.*, vol. 14, no. 2, pp. 406–420, Feb. 2014.
- [75] X. Yang, W. Song, and D. De, "Live Web: A SensorWeb portal for sensing the world in real-time," *Tsinghua Sci. Technol.*, vol. 16, no. 5, pp. 491–504, 2011.
- [76] Y. Fathy, P. Barnaghi, S. Enshaeifar, and R. Tafazolli, "A distributed in-network indexing mechanism for the Internet of Things," in *Proc. IEEE 3rd World Forum Internet Things (WF-IoT)*, 2016, pp. 585–590.
- [77] J. S. Pereira, P. A. Smirnov, J. A. O. Martinez, R. U. Rezvani, P. U. G. Gil, and N. U. Pourshahrokhi, *D4.2 Large Scale IoT Crawling, Indexing and Ranking*, Siemens AG Austria, Vienna, Austria, 2020.
- [78] C. Truong, K. Römer, and K. Chen, "Fuzzy-based sensor search in the Web of Things," in *Proc. 3rd IEEE Int. Conf. Internet Things*, 2012, pp. 127–134.
- [79] R. Mietz, S. Gropp, K. Römer, and D. Pfisterer, "Semantic models for scalable search in the Internet of Things," *J. Sensor Actuator Netw.*, vol. 2, no. 2, pp. 172–195, Mar. 2013, doi: [10.3390/jsan2020172](https://doi.org/10.3390/jsan2020172).
- [80] M. R. Faheem, T. Anees, and M. Hussain, "The Web of Things: Findability taxonomy and challenges," *IEEE Access*, vol. 7, pp. 185028–185041, 2019.
- [81] S. Pattar, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik, "Searching for the IoT resources: Fundamentals, requirements, comprehensive review, and future directions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2101–2132, 3rd Quart., 2018.
- [82] S. De, Y. Zhou, and K. Moessner, "Chapter 1—Ontologies and context modeling for the Web of Things," in *Managing the Web of Things*, Q. Z. Sheng, Y. Qin, L. Yao, and B. Benatallah, Eds. Boston, MA, USA: Morgan Kaufmann, 2017, pp. 3–36. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/B9780128097649000020>
- [83] M. Noura, M. Atiquzzaman, and M. Gaedke, "Interoperability in Internet of Things: Taxonomies and open challenges," *Mobile Netw. Appl.*, vol. 24, no. 3, pp. 796–809, Jun. 2019. [Online]. Available: <https://doi.org/10.1007/s11036-018-1089-9>
- [84] Y. Zhou, S. De, W. Wang, and K. Moessner, "Search techniques for the Web of Things: A taxonomy and survey," *Sensors*, vol. 16, no. 5, p. 600, Apr. 2016, doi: [10.3390/s16050600](https://doi.org/10.3390/s16050600).
- [85] K. Römer, B. Ostermaier, F. Mattern, M. Fahrmaier, and W. Kellerer, "Real-time search for real-world entities: A survey," *Proc. IEEE*, vol. 98, no. 11, pp. 1887–1902, Nov. 2010.
- [86] D. Zhang, L. T. Yang, and H. Huang, "Searching in Internet of Things: Vision and challenges," in *Proc. IEEE 9th Int. Symp. Parallel Distrib. Process. Appl.*, 2011, pp. 201–206.
- [87] F. Choudhury, F. Rahman, R. Jamil, and N. Mansoor, "Sensor searching techniques in Internet of Things: A survey, taxonomy, and challenges," *ICT Anal. Appl.*, vol. 2, no. 154, pp. 755–763, 2020.
- [88] S. Evdokimov, B. Fabian, S. Kunz, and N. Schoenemann, "Comparison of discovery service architectures for the Internet of Things," in *Proc. IEEE Int. Conf. Sensor Netw. Ubiquitous Trustworthy Comput.*, 2010, pp. 237–244.
- [89] C. Manning, P. Raghavan, and H. Schütze, "Introduction to information retrieval," *Nat. Lang. Eng.*, vol. 16, no. 1, pp. 100–103, 2008.
- [90] N. K. Tran, M. A. Babar, Q. Z. Sheng, and J. Grundy, "A framework for Internet of Things search engines engineering," in *Proc. 26th Asia-Pac. Softw. Eng. Conf. (APSEC)*, 2019, pp. 228–235.
- [91] N. K. Tran, "A reference architecture and a software platform for engineering Internet of Things search engines," Ph.D. dissertation, School Comput. Sci., Univ. Adelaide, Adelaide, SA, Australia, Aug. 2018. [Online]. Available: [https://digital.library.adelaide.edu.au/dspace/bitstream/2440/117732/1/Tran2018\\_PhD.pdf](https://digital.library.adelaide.edu.au/dspace/bitstream/2440/117732/1/Tran2018_PhD.pdf)
- [92] W. G. Hatcher, C. Qian, W. Gao, F. Liang, K. Hua, and W. Yu, "Towards efficient and intelligent Internet of Things search engine," *IEEE Access*, vol. 9, pp. 15778–15795, 2021.
- [93] A. Cimmino and R. García-Castro, "WoTHive: Enabling syntactic and semantic discovery in the Web of Things," *Open J. Internet Things*, vol. 8, no. 1, pp. 54–65, 2022. [Online]. Available: [https://www.ronpub.com/ojot/OJIOT\\_2022v8i1n06\\_Cimmino.html](https://www.ronpub.com/ojot/OJIOT_2022v8i1n06_Cimmino.html)
- [94] V. Charpenay and S. Käbisch, "On modeling the physical world as a collection of things: The W3C thing description ontology," in *Proc. Semantic Web*, 2020, pp. 599–615.
- [95] M. M. Rahman, M. Kutlu, T. Elsayed, and M. Lease, "Efficient test collection construction via active learning," in *Proc. ACM SIGIR Int. Conf. Theory Inf. Retrieval (ICTIR)*, 2020, pp. 177–184. [Online]. Available: <https://doi.org/10.1145/3409256.3409837>
- [96] Y. Zhang, X. Liu, and C. Zhai, "Information retrieval evaluation as search simulation: A general formal framework for IR evaluation," in *Proc. ACM SIGIR Int. Conf. Theory Inf. Retrieval (ICTIR)*, 2017, pp. 193–200. [Online]. Available: <https://doi.org/10.1145/3121050.3121070>
- [97] M. Aziez, S. Benharzallah, and H. Bennoui, "Service discovery for the Internet of Things: Comparison study of the approaches," in *Proc. 4th Int. Conf. Control Decis. Inf. Technol. (CoDIT)*, 2017, pp. 599–604.
- [98] K. R. Lavingia and R. Mehta, "Information retrieval and data analytics in Internet of Things: Current perspective, applications and challenges," *Scalable Comput. Pract. Exp.*, vol. 23, no. 1, pp. 23–34, 2022.
- [99] I. Barclay, C. Simpkin, G. Bent, T. L. Porta, D. Millar, and A. Preece, "Enabling discoverable trusted services for highly dynamic decentralized workflows," in *Proc. IEEE/ACM Workflows Support Large-Scale Sci. (WORKS)*, 2020, pp. 41–48.
- [100] X. Liu, X. Yang, Y. Luo, and Q. Zhang, "Verifiable multi-keyword search encryption scheme with anonymous key generation for medical Internet of Things," *IEEE Internet Things J.*, vol. 9, no. 22, pp. 22315–22326, Nov. 2022.
- [101] Y. Bao, W. Qiu, and X. Cheng, "Secure and lightweight fine-grained searchable data sharing for IoT-oriented and cloud-assisted smart healthcare system," *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2513–2526, Feb. 2022.
- [102] K. Khadir, N. Guermouche, T. Monteil, and A. Guittoum, "Towards avatar-based discovery for IoT services using social networking and clustering mechanisms," in *Proc. 16th Int. Conf. Netw. Service Manag. (CNSM)*, 2020, pp. 1–7.
- [103] F. Luis-Ferreira and R. Jardim-Gonçalves, "Modeling of things on the Internet for the search by the human-brain," in *Technol. Innov. Internet Things*, L. M. Camarinha-Matos, S. Tomic, and P. Graça, Eds. Costa de Caparica, Portugal: Springer, 2013, pp. 71–79.



**Cristyan Manta-Caro** received the Electronics Engineering and Master of Science in Information and Telecommunications degrees from the District University of Bogotá-Francisco José de Caldas, Bogotá, Colombia, in 2005 and 2007, respectively, and the Ph.D. degree in information and communications from the University of Granada, Granada, Spain, in 2023.

He is an experienced Researcher of Internet and Web of Things with the Computer Science and AI Department, Research Centre for Information and Communications Technologies, University of Granada. With over 20 years of experience, he is an SME in Automation, Orchestration, and Network Evolution. He leads transformation and evolution programs for digital and telecommunications providers CSPs in Europe, the U.K. and LATAM. His research interests include SDN architectures, future Internet, DevNet and DevOps, information retrieval, cloud technologies for IoT, and the Web of Things.





**Annalina Caputo** is currently pursuing the Laurea degree (Hons.) in computer science from the University of Bari, Bari, Italy, and the Ph.D. degree in computer science from the Department of Computer Science, University of Bari.

She is an Assistant Professor with the School of Computing, Dublin City University, Dublin, Ireland, and she is the Academic Lead for M.Sc. of Artificial Intelligence for ICT Skillnet. She is an Academic Collaborator of the ADAPT Centre, Dublin, where she works on topics related to personalisation, information extraction and retrieval, and natural language processing. Prior to this appointment, she was an EDGE COFUND Marie Skłodowska-Curie Action Research Fellow with the ADAPT Centre, Trinity College Dublin, Dublin, in the domain of Temporal Aware Personalised Information Retrieval. She co-organised three evaluation campaigns, QA4FAQ, NEEL-IT, and DIACR-ITA at EVALITA. She has attended several IR conferences and workshops, such as SIGIR, IIR, CLEF, TPDF, and UMAP.

Ms. Caputo was also the Sustainability Chair at ICMR 2020. She serves regularly as a Reviewer and a Program Committee Member for a number of workshops and national and international conferences, such as ECAI, AAAI, ECIR, SIGIR, LREC, SAC, WWW, ISWC, and CLIC-it.



**Juan M. Fernández-Luna** received the bachelor's degree in computer science from the University of Granada, Granada, Spain, in 1994, and the Ph.D. degree from the University of Granada in 2001, working on a thesis in which several retrieval models based on Bayesian networks for information retrieval were designed.

He is working on a thesis in which several retrieval models based on Bayesian networks for information retrieval were designed. He is currently a Professor with the Computer Science Department, University of Granada, Granada. His main research interests include XML retrieval, working in collaboration with Juan F. Huete and Luis M. de Campos in XML personalization, collaborative IR, recommender systems and learning to rank, areas in which they have published papers in prestigious journals and international conferences and edited special issues.