Tracing the complicated yet still relatively unripe area of the Internet of Things search engine—from concepts, to classification, and open issues.

BY NGUYEN KHOI TRAN, QUAN Z. SHENG, M. ALI BABAR, LINA YAO, WEI EMMA ZHANG, AND SCHAHRAM DUSTDAR

Internet of Things Search Engine

ADVANCEMENTS UNDER THE moniker of the Internet of Things (IoT) allow things to network and become the primary producers of data in the Internet.[14] IoT makes the state and interactions of real-world available to Web applications and information systems with minimal latency and complexity.[25] By enabling massive telemetry and individual addressing of "things," the IoT offers three prominent benefits: spatial and temporal traceability of individual real-world objects for thief prevention, counterfeit product detection and food safety via accessing their pedigree; enabling ambient data collection and analytics for optimizing crop planning, enabling telemedicine and assisted living; and supporting real-time reactive systems such as smart building, automatic logistics and self-driving, networked cars.[11] Realizing these benefits requires the ability to discover and resolve queries for contents in the IoT. Offering these abilities is the responsibility of a class of software

system called the *Internet of Things search engine* (*IoTSE*).

IoTSE is a complicated and relatively immature research topic. The diversity of its solution space is, arguably, a primary challenge hindering its advance. Such diversity manifests itself in terms of the type of operations within an IoTSE instance (for example, discover content, index, and resolve queries), and the types of IoT content on which those operations are applied. Each combination of operation and content type represents a research area within the IoTSE literature with its own set of technical, social, and political issues. For instance, the IoTSE instances that discover and resolve queries on real-time sensing data from IoT-enabled sensors face the challenge of ensuring the "freshness" of data used for processing queries while minimizing the costly operation of pulling the data from sensors. IoTSE instances working with the actuating functionalities of IoT-enabled things, on the other hand, concern more with understanding the semantics of these functionalities. Due to the diversity of the IoTSE solution space and the lack of a shared vision of what IoTSE is and what it does, it is challenging to communicate the problems and the solutions related to this system.

» key insights

■ **Any collection of information is only as useful as its information retrieval mechanism. Yet a comprehensive search engine is precisely the missing component of the Internet of Things (IoT). While some crawlers for IoT devices exist, an advanced IoT Search Engine (IoTSE) that can resolve queries for IoT content and based on IoT content is still beyond the horizon.**

■ **Following an extensive review of the academic research efforts and industrial projects, this article proposes a meta-path model to describe IoTSE and discusses various IoTSE open issues that have emerged in the review.**

■ **The conceptual model presented lays a foundation for the future integration of identified IoTSE visions.**

The lack of such models and constructs for the communication of IoTSE inhibits more extensive research and development efforts that span research communities over an extended time, which are necessary for the advance of the IoTSE. As the existing studies on IoTSE have primarily focused only on technical issues related to a particular "IoTSE operation – IoT content type" combination, and as the existing reviews and surveys on IoTSE have primarily focused on a particular type of IoTSE, the lack of models and constructs to communicate and classify IoTSE, which are applicable to its diverse solution space, has not been addressed in the existing literature.

In this article, we introduce the fundamental concepts related to the functionality of an IoTSE instance and a model called meta-path to provide a comprehensive yet succinct description of IoTSE instances by their functionality. We report a classification of IoTSE instances based on their meta-path description and present the representative IoTSE prototypes in each class. Finally, we discuss several open issues in the IoTSE research and development.

Methodology

The concepts and models presented in this article were generated from a structured and comprehensive study of the existing research works and industrial projects falling under the moniker of IoTSE. Our methodology was inspired by the systematic literature review method.[8] It comprised four phases: detection, selection, extraction, and synthesis (Figure 1). The *Detection phase* involves identifying potentially relevant articles from various academic sources. The *Selection phase* involves selecting a subset of articles that were high quality and relevant to the study. The *Extraction phase* involves extracting raw data relevant to the questions of the study. The *Synthesis phase* involves synthesizing raw data into knowledge to answer questions of the study.

Our method deviated from the systematic review method by using software tools for automation. Particularly, the detection and selection phase employed an in-house developed tool that queried academic data sources (that is, "primary search") and retrieved articles that had been referenced by the articles detected in the primary search (that is, "snowballing"). We performed the primary search on various academic data sources, including the XML dataset of

DBLP, with the Boolean query "search OR discover *and* Internet of Things OR Web of things". We assessed articles that emerged from the primary and snowballing search against the following selection criteria:

‣ Excluding papers that focus exclusively on physical and network layer.

‣ Excluding papers that focus on utilizing the sensing data from the IoT to extend the Web search

‣ Excluding the information retrieval papers that do not address IoT, *unless they are highly referenced by other relevant works.*

In the extraction phase, we extracted from papers the conceptualization, functionality, and internal operations of the reported IoTSE prototypes. Finally, we synthesized the extracted data into the concepts and models reported in this article.

By applying the reported method, we identified over 200 relevant works on IoTSE that span over a decade. Figure 2a compares the changes in the number of IoTSE-related works published and referenced between 2001 and 2016. The number of IoTSE works published each year has been increasing steadily since 2001. From 2009—the birth year of the IoT, this number has risen sharply and peaked at 38 works a year in 2014 and 2015. There was a drop in the number of published works in 2016, which we contribute to the fact that our prima-

ry study selection concluded by the end of that year and therefore missed the accepted-yet-unpublished works.

The changes in the number of referenced IoTSE works, however, have not assumed a similar pattern with the number of published works. Between 2001 and 2010, other works referenced most of the published works at least once. However, over the following six years, this number dropped gradually, and the gap between the number of published and referenced works widened. By 2015, only 13% of published IoTSE works received in-field citations.

Figure 2b compares the number of referenced IoTSE works and the number of in-field citations between 2001 and 2016 to provide more insights into the distribution of *attention* among the IoTSE literature. Despite fluctuations, the number of in-field citations rose steadily from 2001. After peaking at 67 citations in 2010, this number began to drop sharply. From these figures, we can see that a group of 29 works published between 2010 and 2012 received over 45% of the total number of in-field citation. This result might indicate that the perception of what an IoTSE instance is and what it should do have been driven by a subset of IoTSE works. A comprehensive analysis on the IoTSE literature and the internal operations of representative IoTSE prototypes is available elsewhere.[20]
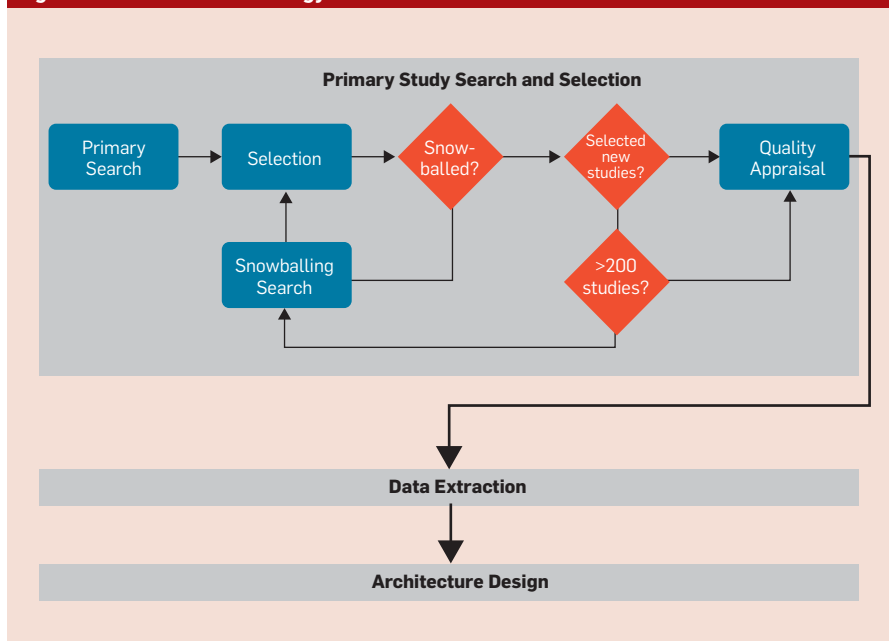
IoTSE Concepts

The Internet of Things comprises IoT *things*—physical objects enhanced with computing and networking capabilities and are potentially accessible via the Internet. For instance, a light bulb equipped with microcontrollers and wireless communication capability is an IoT thing that is commonly found in home automation applications. IoT things offer IoT *content*, such as the digital representation, data records, real-time sensor readings, and functionality that are offered by or related to things.

The IoT content appearing in the IoTSE literature can be organized into four types: representation, static information, dynamic information, and functionality. Figure 3 depicts four IoT content types of an IoT-enabled lightbulb. The representative content of the lightbulb comprises an HTML document that acts as a homepage of the light bulb for interacting with human users, and a JSON document that described the light bulb to machine agents. The dynamic information content of the light bulb denotes either the whole stream of energy consumption readings of the light bulb or the latest value in that stream. Due to the constant update of the lightbulb, these contents are "dynamic". The static information content comprises the archived sensing data, the Web articles related to the lightbulb, and the records of its journey across supply chains. Finally, the functionality content includes *actuating services* that the lightbulb offers to alter its operation (toggling its power, changing its light color).

Discovery activity. An IoTSE instance processes queries on various collections of IoT content. When these collections are not available, an IoTSE instance must carry out the *discovery activity* to detect IoT content in a local or global scope, and optionally collect the content into its internal storage. More than 90% of the assessed IoTSE prototypes include discovery activities.

On a global scale, the content discovery problem can be framed as a Web crawling problem to identify a subset of Websites that serve IoT content (that is, IoT data sources) and retrieve the URI of the IoT content from those sites. In the existing literature, this crawling either relies on human's guidance[16] or the standard compliance of data sourc-

Figure 1. Research methodology.

Primary Study Search and Selection

Primary Search → Selection → Snow-balled? → Selected new studies? → Quality Appraisal

Snowballing Search

>200 studies?

Data Extraction

Architecture Design

es.[12] On a local scale, the content discovery can be addressed as a wireless discovery problem, in which an IoTSE instance either broadcasts beacon signals for things to register themselves, or detects and queries things directly to retrieve their content.[22,26] Local discovery can also be addressed as a service discovery problem in local area networks, using technologies such as multicast Domain Name System (mDNS) or Bonjour.[a] Semantic discovery is an alternative perspective on the content discovery problem. It concerns with detecting the semantics of IoT content and can be addressed by translation content to known data models.[9]

Search activity. The *Search activity* denotes the process of identifying a subset of discovered IoT content as search results of a given query. All assessed IoTSE prototypes covered this activity.

Formally, let c be an item of IoT content, and C be the collection of all content discovered by an IoTSE instance. For each query q, a set of contents c_q that are relevant to the query exists. The task of an IoTSE instance is to construct the result set $\sim c_q$ that approximates the unknown c_q by evaluating the relevance of each IoT resource against the given query with a relevance function $f(c,q)$. If the relevance function produces binary result, the process is considered *selection* or *lookup*:

$Selection: c_q = \{c \in C | f(c,q)=1\}$,
where $f(c,q): C \times Q \rightarrow \{0,1\}$

If the relevance function produces a real value, the result set contains resources whose scores are higher than a predefined threshold α. This process is called *resource scoring*.

$Scoring: r_q = \{c \in C | f(c,q) > \alpha\}$,
where $f(c,q): C \times Q \rightarrow \mathbb{R}$

The selection process cannot determine the degree of relevance of IoT content, and therefore can be considered less advanced compared to the scoring process. However, we discovered that nearly half of the analyzed IoTSE prototypes utilized selection. Most of the remaining prototypes scored IoT content based on its distance from a given query in a multi-dimensional space.
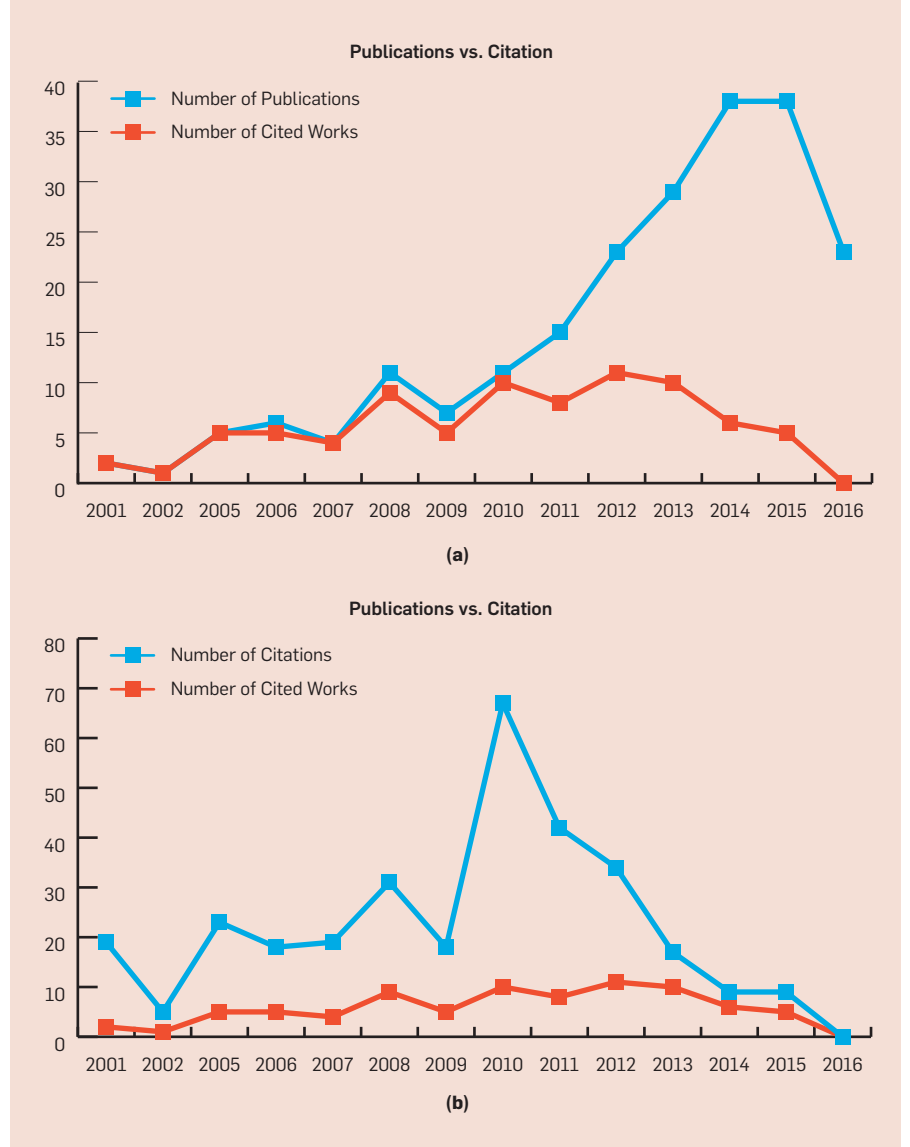
a https://developer.apple.com/bonjour/

The storage and indexing of the discovered IoT content link the discovery activity with the search activity. Most of the existing IoTSE prototypes address the heterogeneity of IoT content by limiting the type and format of content and handle each type independently. For example, IoT-SVK[2] utilizes two B+ trees and an R tree index to address textual description, numeric sensing data, and location of things separately. Some IoTSE prototypes, such as DiscoWoT,[9] address the heterogeneity problem by mapping various formats of IoT content onto a common format for processing.

Meta-path. The lack of a descriptive and comprehensive model to communicate and classify the functionality of IoTSE instances was a major problem

identified from our analysis. For instance, the term "object search" has been used to describe various types of IoTSE instances, which process queries on various types of content—real-time state, description, functionality of things—and return various types of IoT content including sensing data, location, data records, and actuating services of relevant IoT things.
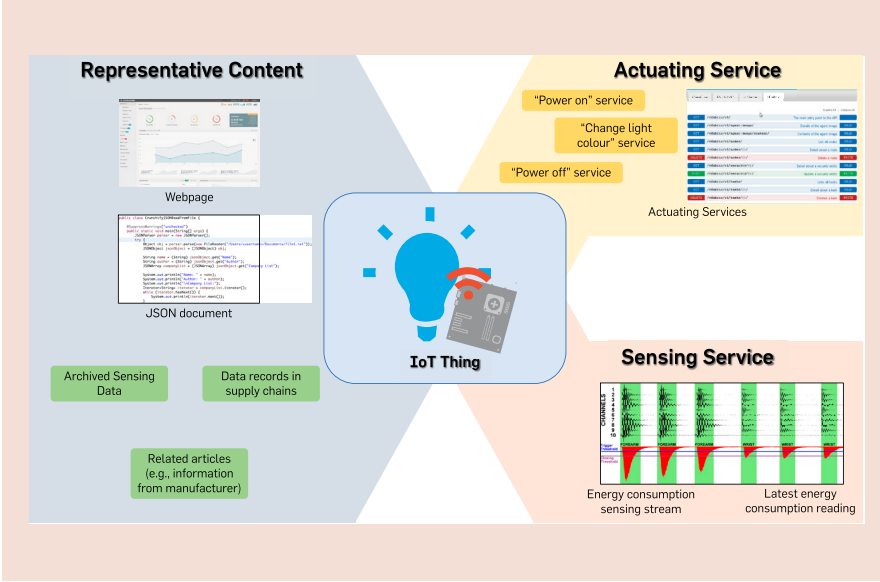
The existing models describe an IoTSE instance either by the type of IoT content that is utilized for processing queries or returned as search results, without considering the relationship between them. Different from the previous types of search engine systems, such as Web search engines, IoTSE can utilize a combination of different IoT

Figure 2. Statistics regarding publication count and number of in-field citations of collected articles.

Publications vs. Citation

(a)

Publications vs. Citation

(b)

Figure 3. Four types of IoT content of an IoT-enabled lightbulb.

Representative Content

Webpage

JSON document

Archived Sensing Data

Data records in supply chains

Related articles (e.g., information from manufacturer)

IoT Thing

Actuating Service

"Power on" service

"Change light colour" service

"Power off" service

Actuating Services

Sensing Service

Energy consumption sensing stream

Latest energy consumption reading

Figure 4. (Upper) Meta-graph from a concrete IoT network, and meta-path $R+D \to T \to D$.

Internet of Things

A: "Power on" service A: "Change light colour" service

A: "Power off" service

R: JSON description

Lightbulb   Located in

R: HTML description

S: Energy consumption   Meeting room

Staff

Uses

R: Homepage

Meta-graph

Actuating Serv.

Representative   Thing   Thing

Thing   Representative

Sensing Serv.

content types to assess a query and to derive search results. Moreover, the types of IoT content appearing in a query influence the internal operations of an IoTSE instance.[20] As a result, an IoTSE model must capture succinctly both the types of involving IoT content and the relationships among those types. Terms such as "object search" are inadequate. To address this issue, we propose a model called *meta-path*.

Before defining meta-path, it would be helpful to introduce the idea of modeling the Internet of Things as a heterogeneous graph, which was inspired by PathSim.[17] The nodes in this graph consist of IoT contents and IoT things that own these contents. The edges

that link things and content denote a possessive relationship between them. The edges that link things denote their possible correlations, such as sharing owners or operation environments.[23,24]

From a concrete graph, we can derive a meta-graph that presents relationships between types of nodes. Each node in a meta-graph is either a type of IoT content or a thing. An edge between a content type and a thing represents the content type is offered by the thing. An edge between two things represents a correlation between them. Different types of thing-thing edges represent different forms of correlation between things. These thing-thing relationships can enable interesting queries such as

finding all employees who had been in the meeting rooms that reported an abnormal energy consumption. However, we have not discovered IoTSE prototypes taking advantage of these correlations. Therefore, for the sake of clarity, we will not depict detailed thing-thing correlations in the following discussions.

Figure 4 depicts the IoT infrastructure in a smart building as a heterogeneous graph (upper) and the derived meta-graph. IoT things in this illustration consist of a smart light bulb, a meeting room, and a staff member who uses these facilities. The light bulb has seven IoT content items of four classes. The meta-graph captures relationships between IoT content types and things, as well as among things. For instance, two representative content items of the lightbulb are captured in the meta-graph as a single link between the representative content type and a thing.
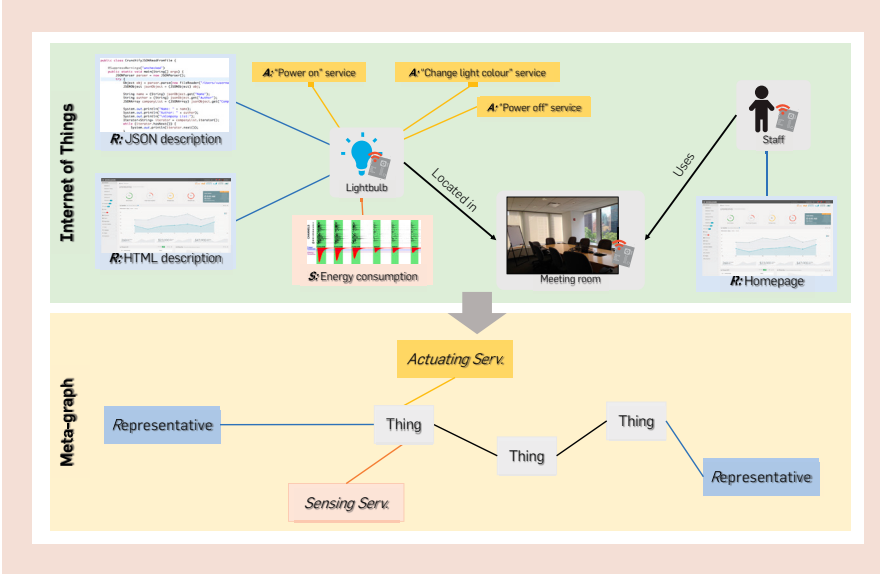
A meta-path is a sequence of edges on the meta-graph from one type of IoT content, through various IoT things, to another type of IoT content. In the IoTSE context, each meta-path can model the relationship between a type of IoT content used for assessing query and a type of IoT content used for deriving search results. By aggregating multiple meta-paths, we can model an IoTSE instances that utilize multiple types of IoT content.

A meta-path can be represented as follows:

$$(Query\ content\ type) \to Things* \to (Result\ content\ type).$$

To demonstrate the meta-path model, we will model an IoTSE instance that queries for "homepages of IoT-enabled light bulbs which are reporting an abnormal energy consumption" as an example. This query can be decomposed into two subqueries: "finding the virtual representative of things, which are lightbulbs" and "finding the virtual representative of things, which are reporting an abnormal energy consumption." The first subquery involves assessing each discovered representative (Representative) to determine whether it belongs to a lightbulb (Thing) and returning the representative of that lightbulb (Representative) as the search result. This subquery can be modeled with the meta-path $R \to T \to R$.

The second subquery involves as-

sessing each discovered sensing data stream (<u>D</u>ynamic IoT content) to detect an abnormality, finding the <u>T</u>hing that offers such data stream, and returning the representative of that thing (<u>Rep</u>resentative) as the search result. This subquery can be modeled with the meta-path $D{\rightarrow}T{\rightarrow}R$.

By aggregating the two subqueries, we can model the IoTSE instance with the aggregated meta-path $R+D{\rightarrow}T{\rightarrow}R$. The IoTSE class addressing this meta-path is the second most common class in the IoTSE literature.

A Meta-path-based Classification System for IoTSE

IoTSE instances can be classified in various dimensions, from implementation technologies[27] to query processing behavior[6,15] and the maturity of their development.[3] Alternatively, we can classify IoTSE instances by their meta-paths, which provide succinct and comprehensive description of their functionality via the type of queries that they support. As mentioned previously, the form of a query that an IoTSE instance addresses influences its internal operations and, therefore, determines its solution space. A meta-path-based classification system will provide insights on what an IoTSE instance is and what it should do, according to the IoTSE literature.

We modeled the IoTSE prototypes selected earlier with the meta-path model and identified eight types of meta-path (Figure 5). Each meta-path represents a class of IoTSE.

$R \rightarrow R$ Class IoTSE

This IoTSE class is the most popular in the literature. Instances of this class resolve queries on ID, metadata, or content of representative IoT content, and return matching representatives as search results. The popularity of this class is a surprising finding, as it does not utilize distinctive content types of IoT, such as sensing data and actuating services, nor the relationship between IoT contents and things.

ForwarDS-IoT[5] is an IoTSE prototype processing queries on semantic description of things, stored in a federation of repositories. Queries in ForwarDS-IoT specify conditions on metadata of IoT things and are translated into SPARQL queries. This prototype supports both synchronous and asynchronous query process-

ing. DiscoWoT[9] is another prevalent prototype belonging to the $R{\rightarrow}R$ class. This system accepts identities of IoT contents as "queries" and returns representation of the given contents in a common format as "search results". Its operation is based on crowd-sourced strategies for translating different types of resource description into the common format. Coverage of strategies represents "discovered IoT content" of DiscoWoT.

$D + R \rightarrow T{\rightarrow}R$ Class IoTSE

This class of IoTSE resolves queries on both sensing data streams (that is, dynamic IoT content) and representatives of IoT things, and returns the representatives of things that satisfy both query criteria. Efficient processing of dynamic IoT content is a primary challenge of this IoTSE class. The following two works represent two prevalent approaches to address this challenge.

IoT-SVK[2] searches for IoT things based on their textual description and real-time sensing values, with respect to spatial and temporal constraints. This search engine collects sensor readings continuously in parallel to the query assessment. To handle the constant influx of sensing data, IoT-SVK utilizes two B+ trees and R tree indexes, which are distributed across a hierarchy of indexing servers. Dyser search engine[12] also searches for IoT things based on their description and real-world states, which are derived from their real-time sensing values. Different from IoT-SVK, Dyser does not collect IoT content. Instead, it contacts things to validate their states for every query. To improve the efficiency of this operation, it predicts sensing

values by assuming the existence of repeating periods in sensing data and ranks things according to this prediction to minimize the number of things to validate.
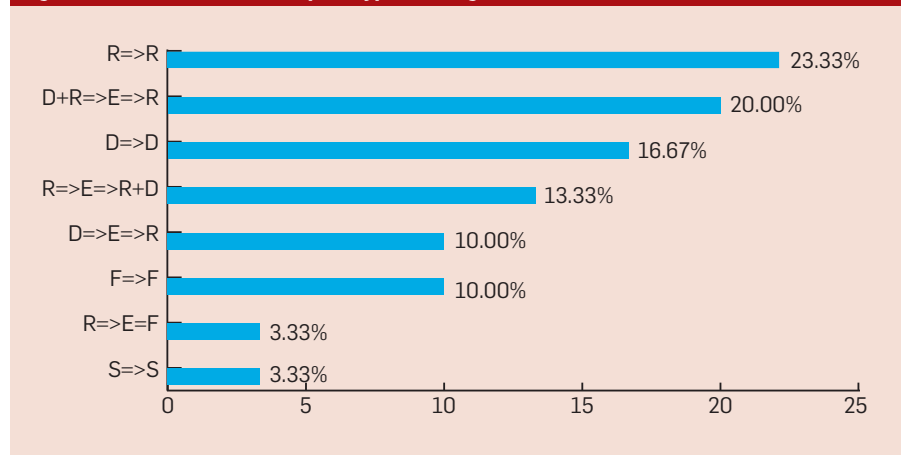
$D \rightarrow D$ Class IoTSE

Search engine instances of this class process queries on metadata and content of sensing data streams and return relevant streams as search results. Key distinction of search engines in this class from the previous one is their focus on low-level sensor readings, instead of high-level states derived from readings.

CASSARAM[13] queries sensing data streams on their contextual information, such as availability, accuracy, reliability and response time. It is motivated by the lack of the search functionality for an increasing number of sensors with overlapping capabilities deployed around the world. CASSARAM utilizes an extension of the Semantic Sensor Network Ontology (SSNO)[1] to describe the contextual information. A user would query this ontology with a SPARQL query generated by the graphical user interface of CASSARAM. This interface also captures the references of the search user. The Euclidian distance between matched sensors and the user reference in a multidimensional space built from different types of sensor contextual information is used for ranking purpose. Top ranked sensing streams are returned as search results.

$R \rightarrow T \rightarrow R + D$ Class IoTSE

This class of IoTSE can be considered as an extension of the class ($R \rightarrow R$). Instances of this class resolve queries

Figure 5. Distribution of meta-path types among reviewed IoTSE works.

Meta-path	Percentage
R=>R	23.33%
D+R=>E=>R	20.00%
D=>D	16.67%
R=>E=>R+D	13.33%
D=>E=>R	10.00%
F=>F	10.00%
R=>E=F	3.33%
S=>S	3.33%

on ID, metadata, or content of representative IoT content, and return both representatives and sensing streams of matching things as search results.

Snoogle[22] is a representative prototype of this IoTSE class. It resolves queries on the textual data stored in IoT things to identify and locate the relevant things. Essentially, Snoogle is a text retrieval system operating on distributed, low-powered repositories. It utilizes a distributed top-k query algorithm with pruning, based on the characteristic of flash memory and Bloom filter, to increase the efficiency of the operation. The representative of matching things, along with their location at the query time (that is, dynamic information IoT content), is returned as search results.

$D \rightarrow T \rightarrow R$ Class IoTSE

This class of IoTSE resolves queries on various aspects of sensing data streams and returns the digital representative of things possessing the matching sensing data streams. Different from class $D+R \rightarrow T \rightarrow R$, search engines in this class do not consider other features of things.

Content-based Sensor Search (CSS)[21] is a representative prototype of this class. It searches for IoT-enabled sensors that produce measurements within a certain range for a certain time prior. CSS contacts sensors to validate their values during query processing instead of collecting IoT content a priori. It utilizes time-independent prediction models (TIPM) to rank sensors based on their probability of having the queried state. These models assume that a sensor reading, which a sensor frequently and continuously reports in the past, has a higher probability to be its current reading. The details of the sensor nodes providing the matching streams are returned as search results.

$F \rightarrow F$ Class IoTSE

This class of IoTSE resolves queries on functionality of IoT things. Considering the popularity of functionality content in real world usage scenarios of IoT (for example, smart home), the limited support for this IoTSE class is a surprising finding. The lack of public datasets and standards for functionality content might have contributed to this limitation.

Mrissa et al.[10] present a search and discovery mechanism for functionalities of physical entities. It aims to discover and expose high-level functionalities of a physical entity that can be realized by a combination of its low-level physical capabilities and functionalities exposed by other entities in the immediate area. These functionalities and capabilities are described in a shared ontology. Each physical entity queries this ontology with a set of SPARQL queries encapsulated in Java functions. This work is part of the avatar architecture from the ASAWoO project.[b]

$R \rightarrow T \rightarrow F$ Class IoTSE

This class of IoTSE resolves queries on the representatives to find relevant things and returns functionality of those things as search results. Kamilaris et al.[7] propose an IoTSE instance that utilizes a DNS-like mechanism to search for IoT things and return their functionalities. These functionalities are presented as RESTful Web services and capable of self-describing with specifications written in the Web Application Description Language (WADL).

$S \rightarrow S$ Class IoTSE

This class of IoTSE resolves queries on static information IoT content. Matching content is returned as search results. Microsearch[18] is an instance of this IoTSE class. It is essentially a downscaled information retrieval system operating on sensor nodes with very limited computing and storage resources. It indexes small textual documents stored in the sensor node and returns the top-k documents that are most relevant to the query terms given by a search user.

Open Issues

As IoTSE research and engineering is a complex and relatively new area, researchers and practitioners face several types of technical challenges. Here, we discuss four open issues, derived from the existing literature, that affect most classes of IoTSE.

Building datasets for IoTSE research. Large-scale, open datasets that contain IoT content, sample queries, and ground truth, are critical to Io-

TSE research. They negate the need for difficult-to-replicate experiments and simplify the experimentation and evaluation of the research works on IoTSE. Research works on IoTSE have utilized some sensing datasets, such as Intel Lab,[c] NOAA,[d] bicycle rental,[e] taxi GPS.[f] Actuating functionality datasets, on the other hand, have not been found in the existing literature. Availability of the sample queries has also been limited, as they tend to be private property of industrial IoTSE instances.[16] Providing access to IoT datasets is a challenge due to their massive size, reaching 21 terabytes a day,[16] and their potential threats to privacy. Given these opportunities and challenges, building open IoT datasets is essential in the IoTSE research.

Ranking IoT contents by their natural order. Natural order ranking denotes the ordering of content by their intrinsic characteristics instead of their relevance to a given query. In large data collections where a massive number of data items can be relevant to a query, a search engine must rely on natural order ranking mechanisms to order and deliver the most relevant search results to its query clients. For example, the ranking of Web pages based on their importance by using link analysis algorithms such as PageRank is a form of natural order ranking.

As the anticipated size of the IoT is even more extensive than the Web, we anticipate that the natural order ranking mechanisms for the IoT content will be an exciting and challenging research topic that will play a crucial role in IoTSE. The first problem in this topic would be defining a natural order that is applicable across different IoT content. For the Web, the level of authority is the natural order of Web pages. For the IoT, what would be the natural order of the heterogeneous IoT content? When this natural order has been defined, the next problem would be developing mechanisms to calculate it on the IoT-scale.

A potential solution to natural order ranking could rely on the quality-

---

of-service metrics of IoT services. Another potential approach could reuse the solution of the Web by constructing a network of hidden links between IoT things[23,24] and applying link analysis algorithms such as Page Rank and its variants to devise a natural ordering of content in the IoT.

Security, privacy, trust. IoTSE instances have the potential to detect and retrieve anything in the IoT, at any place and any time. They bring a wide range of benefits to human users and software agents but also present significant security and privacy risks. IoTSE instances can track a person, monitor an area without consent,[4] and spy into warehouses of competing businesses.[3] Perpetrators can also take advantage of IoTSE to propagate malicious sensing information and actuating services. As future IoT applications might rely solely on IoTSE to acquire IoT content for their operation, misleading information propagated by IoTSE can have severe impacts. For example, by planting sensors that imply a restaurant is full, competitors can drive it out of business. Addressing security, privacy, and trust issues, therefore, is arguably more critical to the success and adoption of IoTSE compared to perfecting its discovery and search algorithms.

Facilitating composition and reuse of IoTSE solution. Across different classes of IoTSE, we have observed shared internal operations such as content discovery, indexing, and searching, albeit with different implementation to serve different types of IoT content. We have also observed the overlaps between various meta-paths, such as between $[D + R \rightarrow T \rightarrow R]$ and $[D \rightarrow T \rightarrow R]$. These observations suggest that prior IoTSE instances can be reused to improve other instances or compose new instances, which might utilize a different meta-path.

Realizing composition and reuse of IoTSE solutions require a common IoTSE architecture and a supporting software infrastructure to support the development, accumulation of IoTSE components, and the engineering of IoTSE instances from those components. Tran et al.[19] propose to utilize a shared software library to facilitate the development of reusable, composable IoTSE components and support

the composition of these components into operational IoTSE instances. Reducing the constraints of the shared library on component developers and simplifying the distribution of components in an IoTSE instance could improve the approach. The Service-oriented Architecture (SOA) is a potential solution to this problem, due to its enforced separation of concern between services and its native support for composition.

Conclusion

Internet of Things Search Engine denotes a software system responsible for discovering and resolving queries on contents of the Internet of Things. Due to the diversity of IoT contents, developing IoTSE is a complex and diverse problem that is still relatively immature. This article introduces concepts, models, and a classification system for IoTSE, which have been generated from a structured and comprehensive study of the literature on IoTSE. We have categorized the latest works into eight classes of IoTSE and presented four major open issues that impact all classes of IoTSE. C

References
1. Compton, M. et al. The SSN ontology of the W3C semantic sensor network incubator group. *Web Semantics: Science, Services and Agents on the World Wide Web 17* (2012), 25–2.
2. Ding, Z., Chen, Z. and Yang, Q. IoT-SVKSearch: A real-time multimodal search engine mechanism for the Internet of Things. *Intern. J. Communication Systems 27*, 6 (2014), 871–897.
3. Evdokimov, S. et al. Comparison of discovery service architectures for the Internet of Things. In *Proceedings of the IEEE 2010 Intern. Conf. on Sensor Networks, Ubiquitous, and Trustworthy Computing.*
4. Frank, C. et al. The sensor internet at work: Locating everyday items using mobile phones. *Pervasive and Mobile Computing 4*, 3 (2008), 421–447.
5. Gomes, P. et al. A federated discovery service for the Internet of Things. In *Proceedings of the 2nd ACM Workshop on Middleware for Context-Aware Applications in the IoT*. Vancouver, Canada, 2015.
6. Jin, X. et al. Where searching will go in Internet of Things? In *Proceedings of Wireless Days*. IFIP, 2011. IEEE.
7. Kamilaris, A., Papakonstantinou, K. and Pitsillides, A. Exploring the use of DNS as a search engine for the Web of Things. *IEEE World Forum on Internet of Things* (WF-IoT). 2014. IEEE.
8. Kitchenham, B.A. and Charters, S.M. Guidelines for performing systematic literature reviews in software engineering. EBSE Technical Report, Ver. 2.3, 2007.
9. Mayer, S. and Guinard, D. An extensible discovery service for smart things. In *Proceedings of the 2nd ACM Intern. Workshop on Web of Things*. San Francisco, CA, 2011.
10. Mrissa, M., Médini, L. and Jamont, J.-P. Semantic discovery and invocation of functionalities for the Web of Things. In *Proceedings of the 23rd IEEE Intern. Conf. on Enabling Technologies: Infrastructure for Collaborative Enterprises*. Parma, Italy, 2014.
11. Ngu, A.H.H. et al. IoT middleware: A survey on issues and enabling technologies. *IEEE Internet of Things Journal 4*, 1 (2017), 1–20.
12. Ostermaier, B. et al. A real-time search engine for the Web of Things. In *Proceedings of the 1st IEEE Intern. Conf. on the Internet of Things*. Tokyo, Japan, 2010.
13. Perera, C. et al. Context-aware sensor search, selection and ranking model for Internet of Things middleware. In *Proceedings of the 14th IEEE Intern. Conf. on Mobile Data Management*. Milan, Italy, 2013.
14. Qin, Y. et al. When things matter: A survey on data-centric Internet of Things. *J. Network and Computer Applications 64* (2016), 137–153.
15. Romer, K. et al. Real-time search for real-world entities: A survey. In *Proceedings of the IEEE 98*, 11 (2010), 1887–1902.
16. Shemshadi, A., Sheng, Q.Z. and Qin, Y. ThingSeek: A crawler and search engine for the Internet of Things. In *Proceedings of the 39th ACM SIGIR Intern. Conf. on Research and Development in Information Retrieval*. Pisa, Italy, 2016.
17. Sun, Y. et al. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. In *Proceedings of the VLDB Endowment 4*, 11 (2011), 992–1003.
18. Tan, C.C. et al. Microsearch: When search engines meet small devices. In *Proceedings of the 6th Intern. Conf. on Pervasive Computing*. Sydney, Australia, 2008. Springer-Verlag Berlin.
19. Tran, N.K. et al. A kernel-based approach to developing adaptable and reusable sensor retrieval systems for the Web of Things. In *Proceedings of the 18th Intern. Conf. on Web Information Systems Engineering*. Moscow, Russia, 2017.
20. Tran, N.K. et al. Searching the Web of Things: State of the art, challenges and solutions. *ACM Computing Surveys 50*, 4 (2017), 1–34.
21. Truong, C. and K. Römer. Content-based sensor search for the Web of Things. In *Proceedings of the IEEE Global Communications Conference*. Atlanta, GA, 2013.
22. Wang, H., C.C. Tan and Q. Li. Snoogle: A search engine for pervasive environments. *IEEE Trans. Parallel and Distributed Systems 21*, 8 (2010), 1188–1202.
23. Yao, L. et al. Things of interest recommendation by leveraging heterogeneous relations in the Internet of Things. *ACM Trans. Internet Technology 16*, 2 (2016), 9.
24. Yao, L. et al. Unveiling underlying connections via mining human-thing interactions in the Web of Things. ACM Trans. *Intelligent Systems and Technology 8*, 5 (2017), 62.
25. Yao, L., Sheng, Q.Z. and Dustdar, S. Web-based management of the Internet of Things. *IEEE Internet Computing 19*, 4 (2015), 60–67.
26. Yap, K.-K., Srinivasan, V. and Motani, M. MAX: Human-centric search of the physical world. In *Proceedings of the 3rd Intern. Conf. on Embedded Networked Sensor Systems*. San Diego, CA, 2005.
27. Zhou, Y. et al. Search techniques for the Web of Things: A taxonomy and survey. *Sensors 16*, 5 (2016), 600.

Nguyen Khoi Tran is a researcher at the University of Adelaide, South Australia.

Quan Z. Sheng is a professor at Macquarie University, Sydney, Australia.

M. Ali Babar is a professor at the University of Adelaide, South Australia.

Lina Yao is a senior lecturer at the University of New South Wales, Sydney, Australia.

Wei Emma Zhang is a lecturer at the University of Adelaide, South Australia.

Schahram Dustdar is a professor at TU Wien, Austria.

Watch the authors discuss this work in the exclusive *Communications* video. https://cacm.acm.org/videos/iot-search-engine