Indonesian Journal of Electrical Engineering and Computer Science
Vol. 28, No. 2, November 2022, pp. 1117~1127
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v28.i2.pp1117-1127

Internet of things search engines: toward a general architecture

Fatima Zahra Fagroud[1], Lahbib Ajallouda[2], El Habib Ben Lahmar[1], Ahmed Zellou[2], Hicham Toumi[3], Sanaa El Filali[1], Youssef Baddi[3]

[1]Laboratory of Information Technology and Modeling Faculty of Sciences Ben M'sik, Hassan II University of Casablanca, BP 7955 Sidi Othman Casablanca, Morocco
[2]Software Project Management Research Team, National School of Computer Science and Systems Analysis, Mohammed V University, Rabat, Morocco
[3]Higher School of Technology-Sidi Bennour, Chouaïb Doukkali University, El Jadida, Morocco

Article Info

Article history:

Received Jan 17, 2022
Revised Aug 20, 2022
Accepted Sep 5, 2022

Keywords:

Architecture
Crawling
Indexing
Internet of things
IoT search engines
Query

ABSTRACT

Internet of things (IoT) represents one of the main data-producing fields on the Internet, given the diversity and growth of smart objects around the world. The advancement of IoT paradigm and the variety of IoT special characteristics present major challenges for IoT search, which attract a great importance by industrials and researchers. Until now, a good deal of research has been focused on the development and implementation of IoT search solutions and tools, though there are still many issues, which must be studied and solved. This paper is interested to IoT search issue and tries to give a guideline to researchers interested in this issue as well as the proposal of a new general architecture for internet of things search engines. The article presents the concept of IoT search engines, a study of various existing solutions and the proposal of a new architecture which is based on 3 components and which respects the various requirements of IoT search engines. The results of this work are prominent as well as they will help researchers to identify future research directions.

Corresponding Author:

Fatima Zahra Fagroud
Laboratory of Information Technology and Modeling, Faculty of Sciences Ben M'sik
Hassan II University of Casablanca
BP 7955 Sidi Othman Casablanca, Morocco
Email: fagroudfatimazahra0512@gmail.com

1. INTRODUCTION

Internet represent the global computer network which make various services available to the public. It has become a great way to communicate, share, work, meet, learn and even trade. Internet of things (IoT) is a new revolution of the Internet. Objects make themselves recognizable and they obtain intelligence by making or enabling context related decisions thanks to the fact that they can communicate information about themselves. They can access information that has been aggregated by other things, or they can be components of complex services [1]. IoT is expected to touch different aspects of the everyday life and behavior of potential users. It will become one of the most vital topics in several industrial and academic sectors [2].

IoT mean system that groups information technology (IT) devices, mechanic and digital equipments, objects, animals or people which are interconnected and can communicate data via a network in a self-sufficient manner, in real time and without any human interaction. It represent a dynamic network infrastructure which include physical and virtual "things" which possess unique identifiers, physical attributes, and virtual characteristics into the information network [3]. IoT things offer IoT content, such as the digital representation, data records, realtime 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 [4].

Search engines are one of the main mechanisms by which users obtain information on the Web. For any individual wanting information about something on the Internet, search engines are the first choice to help find what the user wants [2]. IoT search engine denotes a software system responsible for discovering and resolving queries on contents of the IoT. Due to the diversity of IoT contents, developing IoTSE is a complex and diverse problem that is still relatively immature [4].

Searching in IoT networks has a different goal than the ones that typical search engines adapt where the users would operate the objects locally or remotely. As a result, this distinguishes between both sides and requires a new design concept for an IoT search engine [5]. 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 [4].

This paper aim to provide a brief survey on IoT search issue and to propose a general architecture for IoT search engines. The rest of this paper is organized as follows: In section 2 we will determine: What is Search Engines? and What is IoT Search Engines? Section 3 presents a survey on various solutions for IoT Search. In section 4, we present our proposed general architecture for IoT search engines. Then, in section 5 we will present our results. Lastly, the conclusion.

2. WEB SEARCH ENGINES AND IOT SEARCH ENGINES

2.1. Web search engines

The internet is a popular global information system where users are searching for the relevant information using search engines (SE). The SE is a type of software that organizes the content collected from all across the internet. With SE, users who are wishing to find information only need to enter a keyword about what they had like to see, and the search engine presents the links to the content that resembles what they need. The most popular and widely used SE on the internet is Google where 77 percent of users around the world are using Google Search Engine for searching information on the internet [6].

Search engine is some kind of software, which collects data about web sites. At this point, the collected data includes the web site uniform resource locator (URL), some keywords or keyword groups that define the content of the web site, the code structure that forms the web page and also links provided on the web site. The related collected data is indexed and stored on a database. All of these operations are performed by the search engine software (crawler, spider, bot) [7]. Essential competition factor among search engines is appeared during “relevant result showing, sorting” process. After determining the related pages with the performed querry, they must be shown to users in a sorted list structure. At this point, search engine algorithms take an important role and they try to show the most relevant results for users [7].

The process of search engine is basically based on four sub processes which are:

- Crawling process: This process is based on robots, also called spiders or crawlers that browse the World Wide Web in a methodical and regular manner, to discover new addresses (URLs) judged interesting. They follow the hyperlinks that connect to each other pages [8].
- Indexing process: The objective of this process is to extract keywords considered significant in each page identified by crawlers, and store them in a database, called index database. The purpose of indexing is to facilitate a fast and an accurate information retrieval [8].
- Page ranking process: This process allows ranking and organizing search results, and making the most relevant documents in top of the list [8].
- Query processing process: It’s the process that submits the query and returns the results. Linguistic techniques such as stemming, removal of articles and auxiliary verbs and spelling, are used to reduce noise and increase the recall. In addition, algorithms are applied to identify in the index database, documents that best match the user's query [8].

Briefly, searching robots collect data about each URL and store the collected data in a database. When a user connects to the search engine for a search session, references in the related database are evaluated and obtained results are returned back to user [7].

2.2. IoT search engines

Using web search engines, we can search web pages and explore data on these pages in different topics when we use generic search engine. In the case of search of web pages related to a specific area in the case of using topical search engine. In all cases we cannot search and find physical resources like computer.

IoT has a set of special features, which present great challenges for traditional search engines, in order to respond to these issues and continue the success of search engines with the large number of IoT devices joining the Web every day a new evolution of these tools appeared entitled IoT search engines [9]. IoT search engine is a new type of search tool, which aim to solve the challenges of discovering IoT devices, finding information about IoT, and strengthening security mechanisms. This tool gives relevant results and offer a set of filters to facilitate the specific search and respond to sophisticated request.

When we deeply study the topic of IoT search engines, we can distinguish two major definitions:

- The IoT search can be defined as the process of finding physical objects with the given states captured by the diverse sensors in the IoT. Therefore, the IoT search includes four elements: i) physical objects, ii) diverse sensors that sense the states of physical objects, iii) queries to search for physical objects with a certain state, and vi) a search engine that accepts queries and returns references to objects matching the query [10].
- IoT search engine can also be defined as a specialized search engine that allow their users to find information and learn about internet of things by providing a set of resources that are similar to resources of web search engine.

This tool is useful for the following reasons:

- Automatic resources discovery
- Resources metadata search
- Specific information search
- Find heterogeneous data in real time
- Companies can track the use of their product

The most important requirements of an IoT search engine are:

- real-time retrieval
- spatial-temporal search conditions
- value-based search conditions
- keyword-based search conditions
- historical and recent data are important and should be supported

3. RELATED WORKS

By analyzing a set of works [2], [5], [9], [11]-[29] in the literature related to IoT search engines and IoT resource discovery and control. We discover the existence of an important number of IoT search engines, and we conclude that every search tool has some relative strengths. In this section, we present a summary of some IoT search engines that are developed since 2009.

In 2009 a search engine was launched by John Matterly to crawl the Internet and index discovered services. Today, this search engine entitled shodan is one of the most popular IoT search tools. It is a computer search engine equipped with a graphical user interface that can identify Internet-facing devices and services. Shodan crawls the Internet for available devices and services and stores the collected data, namely internet protocol (IP) address, port, and service banner in a database accessible via http://www.shodanhq.com or via shodan application programming interface (Shodan API). For each discovered service, Shodan scans and stores results repeatedly over time. This yields a time series of results available for each service and accessible to security experts for further processing and analysis [12].

“Snoogle” an information retrieval system (search engine) that follows a two tier architecture. It uses a description compression technique in local registry based on bloom filters and distributed top-k query algorithm to reduce the network overhead. Also, It offers a security and privacy framework for users by leveraging access rights and public key cryptography (PKC) [13].

Faqeeh *et al*. [2] have proposed a new search engine entitled Topical search engine for internet of things which serve to collect all Websites accessible on the Internet than their content related to internet of things. An evaluation of this solution has been realized by submitting various IoT related queries and manually evaluates the relevance of returned pages. The result of this evaluation proves that the proposed search engine is quite fast and provides more precise results than Google, Yahoo and bing.

A multimodal real-time search engine framework (IoT-SVKSearch) [14], [15] is a framework which allow their user to retrieve IoT data (massive and heterogeneous) and supports multimodal search conditions including keyword-based, spatio-temporal, and value-based search. Experiments show that the IoT-SVK search engine performs well in real-time multimodal retrieval of big data from IoT sensors. The system can achieve better performance as the number of index node servers increases, especially for space-time searches.

Censys.io is a search engine designed for IoT search similar to Shodan.io, but it detects only devices connected to IPv4 contrary to shodan which can scan millions of IPv6 addresses by exploiting a loophole within the NTP Pool Project. This tool was proposed by Zakir Durumeric in 2015 to give a response to the following question: what composes the Internet? And query information about hosts and networks. To collect their data, it is based on daily Zmap scans of more than 3 Billion addresses. Censys tags the collected data with security-related properties and device types, allowing easy but powerful search queries through its online search interface and REST application programming interface (API) [16]. As an example application for Censys, the prevalence of the unauthenticated Modbus protocol among supervisory control and data acquisition (SCADA) systems has been studied. Numerous such systems have been found across the globe. However, non-SCADA devices, specifically, the TLS ecosystem for those devices have not been studied [16].

Thingful describes itself as a “discoverable search engine” that provides users with “a unique geographical index of connected objects around the world” (https://thingful.net/). As such, Thingful boasts that it can index across multiple IoT networks and infrastructures. Because this search engine can locate the geographical position of objects and devices, some critics worry that Thingful can also easily access personal data regarding the ways in which users interact and communicate with their connected devices and objects (as in the case of a homeowner communicating with the “intelligent” objects in her smart home) [17].

Datta and Bonnet [18] propose a framework for automatic and efficient discovery of IoT resources in order to allow the discovery of resources, their capacities, their properties and their Uniform resource identifier (URIs), in order to be able to access them independently of the communication technology. The architecture of the framework is made up of 3 layers: Proxy Layer (made up of 2 proxies: Proxy-in dedicated to sensors and Proxy-out dedicated to actuators), Discovery Layer (integrates a search engine which provides the "search" functionality for discovery), Enablement Layer. It is a generic framework that can be deployed in cloud platforms, M2M gateways, and lightweight mobile applications running on smartphones.

COBASEN is a software infrastructure proposed by Lunardi *et al*. [19] which enable scalable search and transparent use of computing devices in IoT environments. It is made up of 2 main components: Context module and search engine.

- Context module: responsible for collecting information related to device context and associated data from the middleware and sending it to the search engine.
- Search Engine: Responsible for indexing device information and responding to queries using the index. The search engine also provides a graphical interface that allows users to select one or more devices, as well as a set of specifications by using specific parameters. Finally, the search engine sends the specification to the middleware layer and gives feedback to the user.

This infrastructure is compatible with any service-oriented architecture (SOA)-based middleware solution that supports the Subscribe/Notify communication model, as well as the data request mode using eXtensible markup language (XML) and web services. A search engine for private IoT networks has been proposed by Habaebi *et al*. [5]. It is a simple, efficient and secure micro search engine that allows users to browse private databases, find and collect data in a useful and secured way. The developed search engine has the capability to produce viable results with acceptable accuracy and minimum delay. However, it is necessary to develop graphical user interface (GUI) because it lacks interactivity and to implement other tests in terms of parallel search. The indexing and ranking algorithms perform well but could be improved tremendously as there is still for improvements to produce more user-oriented results. This search engine applied several security techniques that could be useful to prevent several attacks, however, as a private network, security and privacy need to be one of the main aspects of the system.

Technological partnership between the French search engine Qwant and the Montpellier start-up Kuzzle launches at the innovation and tech fair, which opens in Paris on May 24, 2018. This cooperation aims to optimize the collection and retrieval of data from connected objects. This search engine will be based on the IoT solutions developed by Kuzzle start-up to collect and retrieve information.

Wolfram-Alpha is a computational knowledge discovery engine. It is based on Mathematica that is often known as Wolfram Language (http://www.wolfram.com/mathematica/). The engine allows a user to register their own Internet-connected things such as Twitter, email, Raspberry Pi and others. Each registered device/service has a unique ”databin” and data is updated from each databin every 30 seconds. The main shortcoming is that users can search and query resources if they know their databins. Moreover, Wolfram-Alpha has wolfram data framework (WDF) that summarises and integrates data into a meaningful and expressive form. However, there is no available information about its architecture and technical details [20].

Datta and Bonnet [18] proposed an improved IoTTracker search engine for discovering IoT devices in cyberspace. The authors use IoTTracker to identify IoT devices in the global IPv4 space (discover and identify "stubborn" IoT devices that ARE cannot identify them.). Based on the functionality of the protocol, IoTTracker divides application layer protocols into semi-structured data protocols and unstructured data protocols [21]. For each category, IoTTracker extracts the structure, style or function simhash. Then, IoTTracker uses features pulled from response data to identify IoT devices. Experimental results show that IoTTracker can add 40.76% more identifiable devices compared to ARE (Acquisitional rule-based engine for discovering Internet-of-Things devices [22]).

Another IoT search engine has been proposed in [23] entitled Search Engine for Internet of Things SMPKR IoT search engine which aim to support spatial-temporal-keyword query over mobile object devices. This mechanism typically consists of three different kinds of components: object device (OD), sub-index node (SIN) and primary index node (PIN).

- An object device denotes a sensor or monitoring device attached to a real-world moving object.
- A SIN is in charge of maintaining sensory data gathered from object devices in its vicinity in the form of index trees.
- A PIN has enough storage and computational capacities, and takes charge of collecting data from different SINs in the network [23].

Experimental results show that SMPKR search engine promotes efficient for searching devices with spatial-temporal-keyword constraints compared to chained structure (CS), snoogle (SL) [24] and IoT-SVK (IS) [14], [15]. RT-RCT [9] is a new research tool which has developed in the aim to retrieve connected devices and current available information about these devices in real-time and within minimum delay. The process of this tool start by data collection using network port scanning, then important information extraction and finally present information to user by using a GUI. It's a prominent solution which can be used for security reinforcement, but further improvements must be elaborated like results improvement.

4. RESEARCH METHOD

In Figure 1 we propose a general architecture for IoT search engines. The proposed architecture shows that an IoT search engine can be made up of three modules. These modules are: Crawling module, Indexing module and Query module.

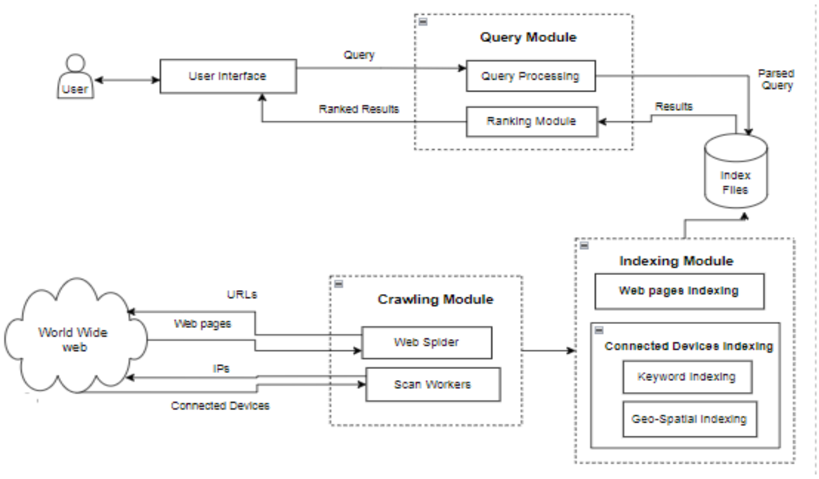

Figure 1. Proposed general architecture

4.1. Crawling module

This module is dedicated to object discovery and data collection. It is made up of two components. These components are: Web spider and scan workers;

- Web spider: this component consists in detecting and analyzing web pages presenting data from connected objects, as well as web pages relating to the topic of the Internet of things. This data collection can be completed either with web crawlers or by a Meta search engine to be developed.
- Scan workers: This component will be responsible for the discovery and detection of connected objects based on the IP address, port number, protocols, etc. This collection of information on connected objects based on IP address, port number, protocols, etc. can be worked out by analyzing the network using the port analysis method or through the flow analysis method. This analysis can be divided over several nodes (Figure 2) in order to collect the maximum amount of information possible as well as in the shortest time.
- These components can be enriched by a subcomponent responsible for updating and optimization due to the dynamism of information such as IP address and device status, update of web pages etc. and the high number of data shows that updating is a crucial element.

- To aggregate the data collected, we can use machine leaning and deep learning algorithms for the classification of connected objects as well as the classification of web pages

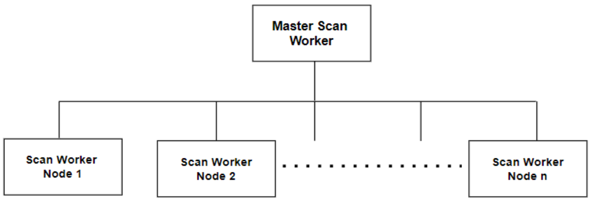

Figure 2. Structure of scan workers

4.2. Indexing Module

Indexing is essential for improving search engine performance and meeting user needs. This is a process making it possible to build a set of "key" elements, making it possible to characterize the data collected and relating to the objects connected in response to a request. Indexing can be: Manual (expert in indexing), Automatic (computer), and Semi-automatic (combination of the two). For that, this module is focused on the indexing of connected objects and web pages (we can use a semantic indexing).

This module will be used to support various types of queries such as keywords and will be based on a two-level tree structure built of a master index and node indexes (Figure 3). It includes two indexing methods:

- Web pages indexing: this method aims to automatically select and extract data from the collected web pages. This means setting up a web wrapper, which will allow the extraction of all the information representative of all the web pages, collected.
- Connected devices indexing: This method aims to build an index structure, which makes it possible to very quickly find and select the connected objects including the requested information. The structure of the index is of the following form: information → {…, IoT device,…} That is to say that each information is mapped to a set of connected objects.
- Given the dynamic behavior of connected objects, the dynamism of information relating to these objects such as IP address and status of devices, updating of web pages etc. and the high number of data demonstrates that the inclusion of an update element updated to this module. This can be done by developing an updating algorithm based on optimization methods and techniques and which takes into consideration all the previously mentioned aspects.
- We can enrich the indexing module with a sub-module that will be linked to the test of attack plans / scenarios (automatic generation of attack scenarios that can affect connected objects). This will allow the identification of the various possible attacks for each type of connected object, which is important for the reinforcement of the security mechanisms of the connected objects.

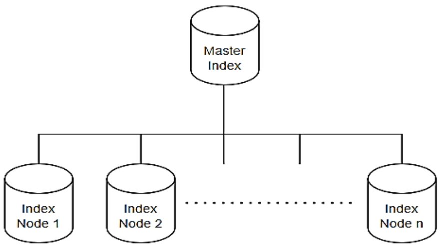

Figure 3. Structure of indexing module

4.3. Query module

This module will be developed on the basis of 2 components:

- Query processing: In an information retrieval system, matching the user's need for information with the relevant answers is the basic problem. A semantically relevant dataset for a query does not necessarily contain the same terms as that query to express the same concepts, especially on the web. In order to better express the user's need for information in a Web context, it is necessary to reformulate his request. The main objective of this element is to translate and enrich the queries expressed by users in natural language into a semantic language like Simple Protocol and RDF Query Language (SPARQL).

- Ranking module: The ranking of the results is done thanks to an algorithm specific to each engine, which uses both logical and mathematical criteria, allowing it to give a score to each page in relation to a given request. The results ranking process makes it possible to process Internet users' queries and make the results the most relevant according to the keywords typed. Several ranking methods exist such as sorting by relevance, sorting by popularity and sorting by audience.

5. RESULTS

5.1. Comparison of web search engines and IoT search engines

Certainly, the difference of the web compared to the Internet of Things in terms of content, dynamicity, links structure as well as scale. In Table 1, we will provide a difference between web search engines and the IoT search engine. This difference has been discussed in many research works such as [10], [30]-[32]. From the results presented in Table 2 we can deduce that the main points of difference between Internet search and IoT search are:

- Search Targets: In internet search, the targets are the information entities, but for IoT search, the targets are broader. IoT search targets can contain information of entities, location and states of the physical objects.
- Interactive Modes: besides the traditional human-computer interaction in cyberspace, the searching in IoT needs the human-computer-object interaction in cyberspace and physical world [10].
- Data Sources: IoT data sources compared with web data sources are vast and heterogeneous because it includes data from the physical world.
- Results Presentation: Unlike the information, presentation was in internet search based on information relevance, in IoT search information is represented based on the state and spatio-temporal information of Things.
- Timeliness: IoT search must be in real time due to the dynamic change and variation of internet of things data.

As well as there is a difference in the search conditions:

- For internet search, the search is based on keywords.
- However, for IoT search, search there are 3 types of search conditions: keyword-based search, spatio-temporal-based search, and value-based search.

Table 1. Differences between the web and the web of things [31]

	Web	IoT
Content Type	Long, unstructured texts (i.e., Web Pages)	Numerical data; Short structured texts
Link Structure	Extensive, explicit link structure between pages (i.e., URL)	Latent links
Dynamicity	Stable; Long lifetime; Slow changing	Volatile; High update rate (up to 1,000,000 per second)
Scale	Over 1 billion Websites	Over 50 billion devices. Interactions happens in local areas

Table 2. Comparison of internet search and Iot search [10]

Characteristic	Internet Search	IoT Search
Search targets	Information entities	Physical objects
Interactive modes	Human + computer	Human + computer + object
Data sources	Cyberspace	Cyberspace + physical world
Results presentation	Information ranking	States + spatial-temporal data
Information timeliness	Not real time	Real time

5.2. Comparison of various IoT search engines

After studying different IoT search engines and their different approaches. We present in Table 3 and Table 4 a comparison of many search engines presented in the literature. These comparatives aim to identify some characteristics of existing search engines and their limits.

Table 3. Properties of existing IoT search engines [32]

Name	Data Source	Purpose	Function
Shodan	IoT devices via crawlers	Assist cyber-security researchers to analyze the influence of vulnerabilities online	Vulnerability search, IoT device meta-data search
Thingful	IoT devices via crawlers and open source data from third-party organizations	Assist users to retrieve information on weather, parking lots, traffic, etc. throughout the world.	Meta-search to different organizations, IoT device meta-data search
Censys	IoT devices via crawlers	Assist cyber-security researchers to analyze the influence of vulnerabilities online	Vulnerability search, IoT device meta-data search
Reposify	IoT devices via crawlers	Assist cyber-security researchers to detect vulnerabilities in organizations	Vulnerability search, IoT device meta-data search

Based on this comparative study we find that:

- Existing IoT search engines update their data periodically, whereas, their results become irrelevant with real-time update of IoT data.
- The exploitation of distributed indexing/multi-index, increase the accuracy of results and can also reduce the response time.
- Displayed results should respect the aspects of privacy and security because it can be a help to hackers.
- The existing IoT search engines work well, but they need to give more importance to the way they present their results so that they can be understood by different users.
- Data collection is a very important step which improve considerably and enrich the response to each query.
- Methodologies and techniques used differ between each of the search engines available.
- Device search engine is the type which occupies more importance

Table 4. Comparison of IoT search engines

	Shodan	IoT-SVKSearch [14], [15]	Topical search engine for IoT [2]	Thingful	Censys	Micro Search Engine [5]	SMPKR IOT search engine [23]
Year	2009	2012	2014	2014	2015	2016	2019
Type of search engine	Devices search engine	Devices search engine	Information search engine	Devices search engine	Devices search engine	Devices search engine	Devices search engine
Network	Wide Area Network	Wide Area Network	-	Wide Area Network	Wide Area Network	Private Network	Wide Area Network
Architecture	Two-tier architecture	Two-tier distributed architecture	Two-tier architecture	N/A	Two-tier architecture	N/A	Three-tier hierarchical searching architecture.
Crawling	Port scanner : SYN Scan	N/A	distributed crawlers	Added by owner of IoT devices	Port scanner : ZMap	MQTT protocol	N/A
Indexing/simple database for storing	Indexing	Indexing	Indexing	Indexing	Indexing	simple database for storing	Indexing
Query	support keyword query	support keyword-based, spatial-temporal, and value-based search conditions	support keyword query	support keyword-spatial query	support keyword query	N/A	support spatial-temporal-keyword query
Performance	Results too numerous to be interpreted effectively and difficult to understand by non-security experts	effective retrieval of massive IoT sampling data	more accurate results than general search engine	accurate results which are presented in a clear way for all users	accurate results which are presented in a clear way for all users	quite strong in terms of searching speed and acceptable results	efficiency of query processing for spatial-temporal-keyword-based object devices
Limits	flexibility in the gathering and presentation of results	Performance of the system depend on the number of index especially in spatial-temporal search	does not process device search and real-time search	includes only some application area of the internet of things	Detect only devices connected in ipv4	No real-time searching Security isn't tested	It's hard to specify the safe region of all IoT devices

6. CONCLUSION

Recently, IoT search engine represent one of the research topics that take a great importance by researchers. Until now many IoT search engines has been presented with the aim of enabling their users to easily and efficiently find and track the connected devices they use, as well as search for information about

IoT and their technologies to learn and understand the subject of IoT and also offers a research tool that gives more precise results. The results generated by existing solutions are difficult to be interpreted and analyzed by non-experts and sometimes irrelevant, which makes it necessary to develop a new IoT search engine that respond to all needs of their users or propose an improvement of existing solutions. In this work, we have reviewed some existing IoT search solutions, and we have proposed a general architecture for IoT search engines. We wish to focus our future work on this topic with the aim of resolving some of their open issues.

REFERENCES

[1] K. K. Patel, S. M. Patel, and P. G. Scholar, "Internet of things-IoT: definition, characteristics, architecture, enabling technologies, application & future challenges," *International Journal of Engineering Science and Computing*, vol. 6, no. 5, pp. 1–10, 2016, doi: 10.4010/2016.1482.

[2] M. Faqeeh, M. Al-Ayyoub, M. Wardat, I. Hmeidi, and Y. Jararweh, "Topical search engine for Internet of Things," in *2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA)*, Nov. 2014, vol. 2014, pp. 829–835, doi: 10.1109/AICCSA.2014.7073287.

[3] F. Z. Fagroud, L. Ajallouda, E. H. Ben Lahmar, H. Toumi, A. Zellou, and S. El Filali, "A brief survey on internet of things (IoT)," in *Lecture Notes in Networks and Systems*, vol. 211 LNNS, 2021, pp. 335–344.

[4] N. K. Tran, Q. Z. Sheng, M. Ali Babar, L. Yao, W. E. Zhang, and S. Dustdar, "Internet of things search engine," *Communications of the ACM*, vol. 62, no. 7, pp. 66–73, Jun. 2019, doi: 10.1145/3284763.

[5] M. H. Habaebi, A. Al-Haddad, A. Zyoud, and G. Hijazi, "Micro Search Engine for IoT: An IoT search engine prototype for private networks," *Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering)*, vol. 11, no. 2, pp. 123–131, Jun. 2018, doi: 10.2174/2352096511666180117144450.

[6] A. Shahzad, D. W. Jacob, N. M. Nawi, H. Mahdin, and M. E. Saputri, "The new trend for search engine optimization, tools and techniques," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 3, p. 1568, Jun. 2020, doi: 10.11591/ijeecs.v18.i3.pp1568-1583.

[7] N. Yalçın and U. Köse, "What is search engine optimization: SEO?," *Procedia - Social and Behavioral Sciences*, vol. 9, pp. 487–493, 2010, doi: 10.1016/j.sbspro.2010.12.185.

[8] I. Saif, A. S. Doukkali, A. Enaanai, and E. H. Benlahmar, "GENAUM: New semantic distributed search engine," *Journal of Mobile Multimedia*, vol. 13, no. 3–4, pp. 210–221, 2017.

[9] F. Z. Fagroud, E. H. Ben Lahmar, H. Toumi, Y. Baddi, and S. El Filali, "Rt-rct: An online tool for real-time retrieval of connected things," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 5, pp. 2804–2810, Oct. 2021, doi: 10.11591/eei.v10i5.2901.

[10] H. Ma and W. Liu, "A progressive search paradigm for the internet of things," *IEEE MultiMedia*, vol. 25, no. 1, pp. 76–86, Jan. 2018, doi: 10.1109/MMUL.2017.265091429.

[11] I. Nadim, Y. El ghayam, and A. Sadiq, "Towards a semantic web of things framework," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 8, no. 4, p. 443, Dec. 2019, doi: 10.11591/ijai.v8.i4.pp443-450.

[12] B. Genge and C. Enăchescu, "ShoVAT: Shodan-based vulnerability assessment tool for Internet-facing services," *Security and Communication Networks*, vol. 9, no. 15, pp. 2696–2714, Oct. 2016, doi: 10.1002/sec.1262.

[13] S. Abdellatif, O. Tibermacine, and A. Bachir, "Service discovery in the internet of things: a survey," in *Lecture Notes in Networks and Systems*, vol. 64, 2019, pp. 60–74.

[14] Z. Ding, Z. Chen, and Q. Yang, "IoT-SVKSearch: A real-time multimodal search engine mechanism for the internet of things," *International Journal of Communication Systems*, vol. 27, no. 6, pp. 871–897, Jun. 2014, doi: 10.1002/dac.2647.

[15] Z. Ding, X. Gao, L. Guo, and Q. Yang, "A hybrid search engine framework for the Internet of Things based on Spatial-Temporal, Value-based, and Keyword-based conditions," in *Proceedings - 2012 IEEE Int. Conf. on Green Computing and Communications, GreenCom 2012, Conf. on Internet of Things, iThings 2012 and Conf. on Cyber, Physical and Social Computing, CPSCom 2012*, Nov. 2012, pp. 17–25, doi: 10.1109/GreenCom.2012.13.

[16] N. Samarasinghe and M. Mannan, "Short paper: TLS ecosystems in networked devices vs. web servers," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10322 LNCS, 2017, pp. 533–541.

[17] H. Tavani, "Search engines and ethics," in *Stanford Encyclopedia of Philosophy*, 2012, pp. 5–6.

[18] S. K. Datta and C. Bonnet, "Search engine based resource discovery framework for Internet of Things," in *2015 IEEE 4th Global Conference on Consumer Electronics (GCCE)*, Oct. 2015, pp. 83–85, doi: 10.1109/GCCE.2015.7398707.

[19] W. T. Lunardi, E. De Matos, R. Tiburski, L. A. Amaral, S. Marczak, and F. Hessel, "Context-based search engine for industrial IoT: Discovery, search, selection, and usage of devices," in *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*, Sep. 2015, vol. 2015-October, pp. 1–8, doi: 10.1109/ETFA.2015.7301477.

[20] Y. Wu, H. Huang, C.-X. Wang, Y. Pan, and Y. Fathy, "Searching for internet-of-things resources: requirements and outlook," in *5G-Enabled Internet of Things*, CRC Press, 2019, pp. 325–337.

[21] X. Wang, Y. Wang, X. Feng, H. Zhu, L. Sun, and Y. Zou, "IoTTracker: An Enhanced Engine for Discovering Internet-of-Thing Devices," in *2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, Jun. 2019, pp. 1–9, doi: 10.1109/WoWMoM.2019.8793012.

[22] X. Feng, Q. Li, H. Wang, and L. Sun, "Acquisitional rule-based engine for discovering Internet-of-Thing devices," in *Proceedings of the 27th USENIX Security Symposium*, 2018, pp. 327–341.

[23] J. Tang, Z. Zhou, L. Shu, and G. Hancke, "SMPKR: Search engine for internet of things," *IEEE Access*, vol. 7, pp. 163615–163625, 2019, doi: 10.1109/ACCESS.2019.2952390.

[24] H. Wang, C. C. Tan, and Qun Li, "Snoogle: A search engine for pervasive environments," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 8, pp. 1188–1202, Aug. 2010, doi: 10.1109/TPDS.2009.145.

[25] C. Kim, Y. Kim, and H. Jung, "IoT task management system using control board," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 1, pp. 155–161, Jan. 2019, doi: 10.11591/ijeecs.v13.i1.pp155-161.

[26] A. G. Shabeeb, A. J. Al-Askery, and Z. M. Nahi, "Remote monitoring of a premature infants incubator," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 3, pp. 1232–1238, Mar. 2019, doi: 10.11591/ijeecs.v17.i3.pp1232-1238.

[27] C. G. Cgseong, J. Y. Kim, and D. J. Park, "Real-time object control system using open source platform," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 1, pp. 313–319, Oct. 2020, doi: 10.11591/ijeecs.v20.i1.pp313-319.
[28] M. A. M. Al-Obaidi, B. J. Ali, J. H.Ssaud, and T. Sutikno, "A hybrid control system for investigating a high level of decision in IoT," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 3, p. 1419, Dec. 2020, doi: 10.11591/ijeecs.v20.i3.pp1419-1426.
[29] A. H. Ali, R. F. Chisab, and M. J. Mnati, "A smart monitoring and controlling for agricultural pumps using LoRa IOT technology," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 1, pp. 286–292, Jan. 2019, doi: 10.11591/ijeecs.v13.i1.pp286-292.
[30] F. Z. Fagroud, E. H. Ben Lahmar, M. Amine, H. Toumi, and S. El Filali, "What does mean search engine for IoT or IoT search engine," in *ACM International Conference Proceeding Series*, Oct. 2019, pp. 1–7, doi: 10.1145/3372938.3372958.
[31] N. K. Tran, "A reference architecture and a software platform for engineering internet of things search engines," Phd. dissertation, The University of Adelaide, 2018.
[32] F. Liang, C. Qian, W. G. Hatcher, and W. Yu, "Search Engine for the Internet of Things: Lessons from Web Search, Vision, and Opportunities," *IEEE Access*, vol. 7, pp. 104673–104691, 2019, doi: 10.1109/ACCESS.2019.2931659.

BIOGRAPHIES OF AUTHORS

Fatima Zahra Fagroud is currently a PHD Student in the Laboratory of Information Technology and Modeling, Hassan II University of Casablanca, where she joined since 2019. She works simultaneously as a professor of computer science at an engineering school in Casablanca since 2021. She received an M.S. in Information Science and Engineering from Faculty of Sciences Ben M'sik in 2018. Her research interests include internet of things, search engines and artificial intelligence. She can be contacted at email: fagroudfatimazahra0512@gmail.com.

Lahbib Ajallouda is a PhD Student at Computer Science and Systems Analysis School (ENSIAS), Mohamed V University, Rabat, Morocco. His research interests are primarily in the area of internet of things, search engines, cloud computing, and machine learning, where he is the author/co-author of over 9 research publications. He can be contacted at email: lahbib_ajallouda@um5.ac.ma.

Prof. El Habib Ben Lahmar is currently a full Professor of computer science in the Department of Mathematics and Computer Science at Faculty of Science Ben M'Sik, Hassan II University of Casablanca, where he has been since 2008. He received his Ph.D. in Computer Science from the ENSIAS in 2007. His research interests span both Web semantic, NLP, Mobile platforms and Data science. He can be contacted at email: h.benlahmer@gmail.com.

Prof. Ahmed Zellou received his Ph.D. in Applied Sciences at the Mohammedia School of Engineers, Mohammed V University, Rabat, Morocco 2008. He is currently a coordinator of the Web Engineering & Mobile Computing branch at ENSIAS Mohamed V university in Rabat, Morocco. His research interests include mediation systems, semantic web, recommendation system, data quality where he is the author/co-author of over 72 research publications. He can be contacted by email: ahmed.zellou@um5.ac.ma.

Prof. Hicham Toumi is a Professor in Higher School of Technology-Sidi Bennour Chouaïb Doukkali University El Jadida, where he has been since 2019. He received a Master degree in network and communication from faculty of science Chouaïb Doukkali University El Jadida in 2013. He obtained a PHD degree in the field of networks security in cloud computing using mobile agents approach at Faculty of science Ben M'Sik, Hassan II University in 2017. Their research interests include cloud computing, security, and internet of things. He can be contacted at email: toumi.doc@gmail.com.

Prof. Sanaa El Filali is currently a full Professor of computer science in the Department of Mathematics and Computer Science at Faculty of Science Ben M'Sik, Hassan II University of Casablanca, where she has been since 2007. She received his Ph.D. in Computer Science from the Faculty of Science Ben M'Sik in 2006. Her research interests include Computer Training, Internet of Things and Information Processing, and much more. She can be contacted at email: elfilalis@gmail.com.

Prof. Youssef Baddi is a full-time Assistant Professor at Chouaïb Doukkali University UCD EL Jadida, Morocco. Phd Thesis degree in computer science from ENSIAS School, University Mohammed V Souissi of Rabat, Morocco, since 2016. He also holds a Research Master's degree in networking obtained in 2010 from the High National School for Computer Science and Systems Analysis-ENSIAS-Morocco-Rabat. His research interests include Information Security and Privacy, the Internet of Things, Networks Security, Software-defined Network, Software-defined Security, IPv6 and Mobile IP. He can be contacted at email: baddi.y@ucd.ac.ma.