Semantic Architecture for Modelling and Reasoning IoT Data Resources based on SPARK

Ahmed Salama¹, Masoud E. Shaheen², Haytham Al-Feet³
Information Systems Department
Faculty of Computers and Information
Fayoum University, Fayoum, Egypt

Abstract—Electronic Internet-of-Things is one of the foremost valuable techniques today. Through it, everything everywhere the globe became connected and intelligent, eliminating the wants to human-to-human interaction to perform tasks. This by changing all of those objects like humans, machines, devices and something around to be simply an internet Protocol (IP) to be expressed within the network environment through completely different sensors and actuators devices which might facilitate the interaction between all of them. These different types of sensors generate a large volume of various information and data. This type of sensor information created it generally useless because of the heterogeneity and lack of interoperability of it that represents it in unstructured form. So, investing from semantic internet techniques might handle these main challenges that face the IoT applications. Hence, the main contribution behind this research aims to boost the performance and quality of sensors information retrieved from IoT resources and applications by using semantic web technologies to resolve the matter of heterogeneity and interoperability and then convert the unstructured sensor data to structured form to realize the next level of investing of sensors employed in IoT applications. Also, the aim through this research to improve the performance of the tremendous amount of information that represents the demonstrated IoT information utilizing Big Data techniques such as Spark and its query language that's named SPARK-SQL as a streaming inquiry language for a colossal amount of information. The proposed architecture demonstrated that utilizing the semantic techniques to model the streaming sensors data improve the value of information and permit us to gather unused information. Moreover, the improvement by using SPARK leads to extend the performance of utilizing this sensor information in terms of the time retrieval of running queries, particularly when running the same queries utilizing the conventional SPARQL inquiry language.

Keywords—Big Data; Internet of Things; Semantic Modelling; Semantic Reasoning; Semantic Rules; Sensors; Apache SPARK; SPARK_SQL

I. INTRODUCTION

Internet-of-Things is considered one of the hottest trends that formulate the progress of information technology development sector. Connecting every object via Internet Protocol (IP) facilitates the intercommunication between human users and machines in different aspects. In this context, there were various researches that focused on the physical side of the IoT applications without representing the importance of the information that is gathered from the resources of Internet of Things devices On this context, there have been various

researches that centered at the physical aspect of the IoT packages without representing the importance of the information which might be collected from the sources of Internet of things devices. IoT is divided into four architectural layers which started with the specified networked things, consisting of wireless sensors and actuators as layer 1, and layer 2 represents each structure of sensors data aggregation and virtual data conversion. Additionally, layer three overviews the role of IT structures in appearing preprocessing of information earlier than it saved into the storage repository. Finally, the extracted information is analyzed, controlled, and loaded directly to the conventional lower back-give up storage systems as shown in Fig. 1.

Hence, the aim of this research is to:

- Build a semantic modeled architecture. This proposed architecture could model the different information fetched from the IoT sensors and actuators such as humidity, temperature, and pressure. This could enrich the meaning of this data and solve the main issue of heterogeneity.
- Build a reasoner tool based on the Description Logic (DL) as one of the Artificial Intelligence languages that depend on semantic web technologies to infer a set of new rules based on a set of existing concepts and individuals after modeling this fetched information.
- Providing the proposed model with the SPARK ecosystem as a big data platform based on Hadoop. This enhancement will increase the performance of the queries performed semantically against the SPARQL query language. This enhancement will illustrate the strength factor that advantages the contribution to others.

The rest of the research is organized as follows: Section 2 presents the literature review that relates to the proposed works. Also, the background technologies which are used through the work are explained in Section 3. In addition to that, the proposed architecture is discussed in Section 4. On the other hand, the implementation processes of the work is presented through Section 5. Also, the results and the comparative study are presented in Section 6. In addition, evaluating the proposed architecture is explained through Section 7. Eventually, Section 8 concludes the paper and discusses the possible directions for future work.
Through this section, we will concentrate on the onset of the most significant logical inquiries that show the
significance of incorporating semantic web advancements with Web of-Things applications. The author of [1]
concentrated on a methodology that supplements the
depictions of Web of-Things assets with progressively nitty-gritty data separated from the demonstrated semantics of
ontologies’ ideas to improve their use and interoperability.
Then again, there is another commitment centered on the
portrayal models for sensor's information utilizing reasonable
ontologies OntoSensor [2] that fabricate a cosmology based
particular model for sensors by excerpting portions of
SensorML [3] depictions. Also, the work presented in [4]
presented a Sensor observation application via semantic web
methodologies, which is called SemSOS, which enables users
to perform sophisticated queries on information from
the environment and data gathered from sensors. While the
community of the World Wide Web Consortium (W3C)
published one of the standard ontologies that integrate with
IoT resources standard ontology named Semantic Sensor
Network (SSN) [5] which describes sensors and their gathered
data. Additionally, the SSN philosophy could deal with the
heterogeneity issue of sensors when assembling their related
information, however, it has a set of confinements in taking
care of the transient or spatial information of sensors assets.
Lamentably, the majority of the current IoT or sensor related
ontologies speak to IoT gadgets just halfway for example
detecting gadgets in SSN cosmology. Which is required for
speaking to more extravagant data identified with IoT
elements and their properties adjusts additionally with one of
the difficulties of Semantic Web research as far as smart
entities, detailed by Sabou M. in [6] for example the portrayal
of a variety of data with respect to smart objects on the IoT.
On the other hand, the IoT-A [7] considers a portion of the
undertakings that stretch out the SSN ontology to simulate
other IoT resources and services. But unfortunately, the IoT-A
model appears to be increasingly unpredictable particularly for
quick client adjustment and responsive conditions. In the same
context, authors in [8] tried to examine how the IoT methods
can be demonstrated utilizing web ontologies that enable them
to straightforwardly convey the strategy usage. On the other
hand, authors in [9] propose an IoT-Lite ontology, by
launching of the semantic sensor organize (SSN) ontology to
depict key IoT ideas permitting interoperability and revelation
of tactile information in heterogeneous IoT stages by a
lightweight semantics. In addition to that, authors In [10]
established a more concluded method for collecting sensor
information by using SASML (Sensors Annotation and
Semantic Mapping Language) to annotate the corresponding
relationship between the SSN ontology and its sensor data in
the mapping file using the RDF Mapping Data Sensor
(SDRM) algorithm.

In addition, the writers in [11] transmitted a major dataset
of sensor metadata and measurements based on a set of
measurements and observations standards that mapped
semantic format such as the Resource Description Framework
(RDF). Also, processing and handling the fetched sensors
information that is stored into an ontology is one of the most
important relevant achieves described in [12]. On the other
hand, handling the huge amount of data semantically is issued
and handled by using big data techniques such as Hive and
Shark as discussed in [13].

Semantic Web is considered one of the Knowledge
Representation (KR) technologies that give plausibility to a
better comprehension of encompassing situations. With a
developing number of sensors and devices linked with the
Internet, semantics play an ever-increasing number of basic
roles as far as information fusion, interoperability, and
understanding. It focuses on how to model different data types
to be processed instead of presenting them only. In addition to
that, the Semantic Web architecture has set basic languages
such as XML, Resource Description Framework (RDF) and
Web Ontology Language (OWL). The ontology can be
considered as a Knowledge Organization and data modeling
tool. Semantic Reasoning is considered the process of
generating new inferences from a collection of given
propositions condition. It is deeply related to the logical
ontologies perspective provided such as OWL Description
Logics (DL) [14].

The Description Logics are used as one of Knowledge
Representation Languages (KR) that depends on the artificial
intelligence tools and Semantic Web technologies to dedicate
a piece of new information based on the given and relevant

Fig. 1. IoT layered Architecture.
concepts of the terminological knowledge of applications. It also provides a logical shape Semantic Web is considered one of the Knowledge Representation (KR) technologies that give plausibility to a better comprehension of encompassing situations. With a developing number of sensors and devices linked with the Internet, semantics play an ever-increasing number of basic roles as far as information fusion, interoperability, and understanding. It focuses on how to model different data types to be processed instead of presenting them only.

In addition to that, the Semantic Web architecture has set basic languages such as XML, Resource Description Framework (RDF) and Web Ontology Language (OWL). The ontology can be considered as a Knowledge Organization and data modeling tool. Semantic Reasoning is considered the process of generating new inferences from a collection of given propositions condition. It is deeply related to the logical ontologies perspective provided such as OWL Description Logics (DL) [14]. The Description Logics are used as one of Knowledge Representation Languages (KR) that depends on the artificial intelligence tools and Semantic Web technologies to dedicate a piece of new information based on the given and relevant concepts of the terminological knowledge of applications. It also provides a logical shape for ontologies of Semantic Web.

In this context, semantic reasoner tools can be used to model the sensors’ data within an ontology environment to infer new information which enriches the given pure and obvious data [15]. This reasoning techniques are formed into Rule-based layer which is defined on the top of the new Semantic Web architecture where different rule languages designed for handling the Semantic Web reasoning tasks such as the Semantic Web Rule Language (SWRL), REVERSE Rule Markup Language (R2ML), RuleML (Rule Markup Language), and Rule Interchange Format (RIF) [16]. On the other hand, leveraging from the huge amount of data collected from sensors is one of the main important challenges which could be enhanced through utilizing big data handling methods and techniques such as Hadoop and SPARK as big data ecosystems that could improve the performance of the retrieved data.

IV. PROPOSED WORK DISCUSSION

The aim behind this investigation is to construct a new semantic architecture to model different data retrieved from sensors systems that control the Internet-of-Things resources and applications. This new architecture aims to handle the issues of information heterogeneity that characterizes the IoT recourses, particularly when utilizing diverse sensor information for diverse utilization purposes. Moreover, we aim through the proposed demonstrate to use as it were the standard substances and ideas of the SSN ontology and after that give the fundamental ontology with the particular and required concepts and properties which are required for abstracting sensors’ data of IoT applications for smart homes.

Hence, the model will enable us to use these different sensors data in different aspects without affecting the meaning or usability of each of them. Also, we are aiming through this research to improve this modeled ontology by extracting more knowledge and information from the gathered data of sensors. This could be performed by using the reasoning language such as Description Logics (DL) and Semantic Web Rule Language (SWRL) to infer new relations and rules that could enrich the value of the proposed model of IoT architecture. The new proposed architecture is organized into four main layers as follow as shown in Fig. 2:

- The First Layer represents the data collection phase in which we gather the data of different types of IoT sensors in real-time such as lighting, temperature, and air conditioner through the WI-FI protocols of IoT such as Arduino Node MCU as a preprocessing layer of the architecture.

- The Second Layer shows the data processing layer in which we model the collected data of sensors using semantic web modeling techniques such as Ontology Engineering languages such as OWL-FULL, OWL-DL, and mapping tools such as RQ2 and D2R to map the sensor data to the semantic format to classify the different sensor data into relevant categories.

- The Third Layer focuses on how to enrich the meaning of the modeled sensor data via reasoning techniques such as Description Logics and SRWL language that process the given sensor data and then infer new relations between them to increase the robustness and homogeneity of data gathered from different sensors types.

- Eventually, the fourth and top layer represents the solution for handling the expected huge amount of sensor data when running the architecture on a large scale. This layer shows the mapping process of the modeled Sensor data to be stored on the Hadoop Distributed File system (HDFS) to be handled and processed by using big data techniques such as Apache SPARK and its query language that is named SPARK-SQL. This enables users to perform different sophisticated queries on the new inferred relations between different modeled sensor data in the proposed ontology which helps researchers to generate accurate statistics and reports regarding understandable, consistent and homogeneous IoT sensors' data by using SPARK-SQL query language to enhance the performance of data retrieval time.

We divide the development process of the proposed architecture into a set of phases as mention before. Through this phase, we used different programming technologies to develop it such as python, java, android, firebase as a mobile database management system and Semantic web tools for modeling and reasoning the IoT resources data. We used the IoT infrastructure techniques such as NodeMCU wireless module for configuring the IoT sensors and actuators Light Dependent Resistor (LDR) sensor for detecting the light mode, lm35 temperature sensor, GPS location sensor, and IR Transmitter sensor, all of these are developed across Arduino environment with integration of python programming language. Also, we use an ontology, RDF, Description Logics
Firstly, we performed the data gathering step by fetching the used sensors data such as different values of temperatures, lighting modes, air conditioners degrees, location latitude and longitude, type of sensors and then stored them into real-time database storage such as Firebase to update them simultaneously.

Secondly, the stored sensor data is then processed by mapping them from Relational Database into the proposed ontology as a real-time database using R2RML as a Mapping Language and RDFLib [19] as python package for asserting the mapped RDF graphs in the ontology. This can be done beside the integration with the standard SSN ontology to model the stored data into the Resource Description Framework (RDF) Format as a traditional framework for representing data in semantic format which help in solving the problem of heterogeneity and give the ability for constructing new relationships between different IoT-Resources to interact together efficiently producing the main objective of this research; the semantic model of IoT sensors' data as shown in Fig. 3 and Fig. 4.

After that, we used DL as an Artificial Intelligence knowledge Representation techniques to build a semantic web reasoning model as well as the SWRL Rule Language to enrich the meaning of ontology by generating a set of new inference rules based on the modeled data for both sensors resources actuators through the proposed semantic ontology and other integrated standard ontologies as shown in Fig. 5.

The work is implemented based on real data of a smart home that has different resources with different sensor’s data. Where the above-mentioned Fig. 3 presented the hierarchical model of the proposed ontology modeled for the IoT domain in the case study. In addition to that, the set of reasoning rules was performed to infer new relations to enrich the meaning of the modeled ontology. Where a new semantic rule is defined for users who use IoT applications that depend on the modeled sensors data in their homes, where sensors detect the temperature of it. This rule could infer a piece of new information for the cases of low-temperature degrees, the sensors of air conditioners should detect off status, otherwise, it shows detect ON status as shown in Fig. 6.

On the other hand, we used the DL techniques that increase the meaning, integrating, and maintaining of the proposed sensors ontology and improve its knowledge representation to handle the issues raised in the traditional reasoning models as shown in Fig. 7.

It named also ALCQIO which provides the modeled ontology with only negation, conjunction, disjunction, and universal and existential restrictions which is called ALC. In addition to that it provides set of additional constructors that refer to number restrictions (Q); inverse roles (I), and nominals (O) as shown in Table I.

| Reasoning Modelled ontology | SWRL | DL |
|------------------------------|------|----|
| Conceptual Sensor Ontology  |      |    |
| Mapping Streaming Sensor Data into RDF | | |
| Reasoning Modelled Ontology  |      |    |
| Running SPARK – SQL Queries  |      |    |
| Running Apache SPARK         |      |    |
| Data over HDFS               |      |    |
| D2R Mapper from RDF to Relational DB | | |

Fig. 2. Proposed Semantic Modelling and Reasoning Architecture.
Fig. 3. Snap of Our Proposed Ontology in OWL DL Format.

Fig. 4. Our Proposed Modelled Sensor Ontology.
during designing and developing the IoT applications to be more efficient for their customers as shown in Fig. 8. On the other hand, we enhance the proposed architecture of the ontological model by applying the most recent techniques that handle the huge amount of data that are expected to be collected especially when applying this ontology on different aspects of IoT applications. Further, the Semantic model of IoT sensor data is formatted into RDF format which facilitates performing SPARQL queries on structure form of Subjects, Predicates, and Objects. Hence, When we propose to enhance this model with the big data techniques such as SPARK, we firstly map the RDF form of the modeled sensor’ data into Relational Format which is the basic form of the SPARK-SQL query language, which is built on the top of Apache Spark. After mapping this data into Relational format, we store the data into the Hadoop Distributed File System (HDFS) to generate a dump of data over Hadoop.

Finally, we performed set of different Semantic queries such as the mentioned query by using SPARK SQL [13] as shown in Fig. 9 query language instead of SPARQL to measure the rate of retrieve time of sensor data which refer to the better performance of running through SPARK-SQL as shown in Fig. 10 as a big data query language against the traditional SPARQL query language as shown in Table II that illustrate the retrieval time that results from performing the same query using the SPARQL and SPARK-SQL that reflect the better performance of SPARQL-SQL against SPARQL.

![Fig. 5. Snap Logical view from our Proposed Integrated Modelled Ontology.](image56x335 to 285x390)

![Fig. 6. SWRL Rule Reasoning Example.](image58x414 to 282x632)

![Fig. 7. Using DL to Create New Semantic Rule Example.](image67x656 to 273x738)

| Symbol | Description |
|--------|-------------|
| ⊓      | empty concept |
| ⋁      | intersection or conjunction of concepts |
| ¬      | negation or complement of concepts |
| ☐      | universal restriction |
| ⋀      | existential restriction |
| ≡      | Concept equivalence |

**TABLE I. THE SYNTAX OF DL (ALCQIO) REASONING**

| Symbol | Description |
|--------|-------------|
| ⊓      | empty concept |
| ⋁      | intersection or conjunction of concepts |
| ¬      | negation or complement of concepts |
| ☐      | universal restriction |
| ⋀      | existential restriction |
| ≡      | Concept equivalence |

![Fig. 8. SPARQL QUERY on Temp Degrees of Locations where it's Air Conditioner observed OFF in our Case Study.](image316x99 to 568x241)

![Fig. 9. The Same Query Performed using SPARK SQL Query Language.](image317x99 to 568x241)
VI. ARCHITECTURE EVALUATION

Through this section, the proposed architecture had to be evaluated based on a set of evaluation criteria defined in [20] such as availability, accuracy, robustness, upgradeability, clearly defined functional layers, and Interoperability as shown in Table III.

VII. CONCLUSION AND FUTURE WORKS

Through this research, a new architecture to model the internet of things using Semantic Web techniques has been designed. We aimed through this research to tackle the problems of heterogeneity and lack of interoperability of data by modeling the data of all the devices and the resources of IoT in a new proposed ontology as well as the standard SSN ontology. In addition to that, we used the semantic Web rules techniques and Description Logics language, to build a reasoner system to enrich the meaning of the proposed modeled ontology and hence enable users to use the modeled sensors data into different aspects. After building this smarter modeled ontology, the sensor’s data is enriched, the data homogeneity and interoperability is enhanced and increased. On the other hand, we enhanced the performance of the proposed architecture, especially when applying it on a huge dataset of IoT sensors that need to be handled using Big Data techniques.

Hence, through the work, we performed a set of sophisticated queries using both SPARQL and SPARQL-SQL as a big data query language on the top of Apache SPARK. These experiments reflect the higher performance of query processing of SPARQL-SQL against SPARQL on the sensor’s data. Hence after performing this enhancement, the performance of executing different sophisticated queries is doubled using SPARQL-SQL instead of using the semantic SPARQL query language. Eventually, the future works will be on how to use the proposed model of ontological IoT to build an expert system that helps users who have these IoT applications to control their applications through their social networks accounts automatically based on the immediate behavior on social media. This will help users to get better and more flexible access to the Internet of Things recourses.

REFERENCES

[1] S. Zander, N. Merkle, and M. Frank, "Enhancing the Utilization of IoT Devices Using Ontological Semantics and Reasoning," Procedia Computer Science, vol. 98, pp. 87-90, 2016.

[2] D. J. Russomanno, C. Kothari and a. O. Thomas, "Sensor ontologies: from shallow to deep models," in Proceedings of the Thirty-Seventh Southeastern Symposium on System Theory, Los Alamitos, CA, USA, 2005.

[3] O. Specification, "OpenGIS Sensor Model Language (SensorML) Implementation Specification," 2007.

[4] C. A. Henson, J. K. Pschorr, A. P. Sheth and K. Thirunarayan, “SensOS: Semantic Sensor Observation Service,” in International Symposium on Collaborative Technologies and Systems, Baltimore, MD, 2009.

[5] M. Compton, P. Barghchi, L. Bermudez, R. Garcia-Castro, O. Corcho, S. Cox, J. Graybeal, and M. Hauswirth, “The SSN ontology of the W3C semantic sensor network incubator group,” Web Semant.Sci. Ser. Agents World Wide Web, vol. 17, p. 25–32, 2012.

[6] K. Kotis and A. Katsanos, “Semantic interoperability on the internet of things: The semantic smart gateway framework,” International Journal of Distributed Systems and Technologies (IJDST), vol. 4, no. 3, pp. pp.47-69, 2013.

[7] W. Wang, S. De, R. Toenjes, E. Reetz, and Klaas, "In Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference," 2012.

[8] V. Caballero, S. Valbuena, D. Vernet, and A. Zaballo, "Ontology-Defined Middleware for the Internet of Things Architectures.," Sensors, vol. 19, no. 5, p. 1163, 2019.
M. e. a. Bermudez-Edo, "IoT-Lite: a lightweight semantic model for the Internet of Things.," in Inl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data C, 2016.

"A Method for Mapping Sensor Data to SSN Ontology," International Journal of u- and e-Service, Science and Technology, vol. 8, no. 6, pp. 303-316, 2015.

M.Alexandra and M.Dunja, "A Framework for Semantic Enrichment of Sensor Data," Journal of Computing and Information Technology, vol. 3, p. 167–173, 2012.

D. F. Barbieri, D. Braga, S. Ceri, E. D. Valle, and M. Grossniklaus, "Querying RDF Streams with C-SPARQL," SIGMOD Record, vol. 39, no. 1, p. 20–26, 2010.

A. Salama, H. Alfeel and H. Mokhtar, “Bridging the gap for retrieving DBpedia data.”, In 2015 Fourth International Conference on e-Technologies and Networks for Development (ICeND), Lodz, Poland, 2015.

L. Li, C. Zhou, J. He, J. Wang, X. Li and X. Wu, "Collective semantic behavior extraction in social networks," Journal of Computational Science- Elsevier, vol. 28, pp. 236-244, 2018.

A. Basu, "Semantic Web, Ontology, and Linked Data," in Web Services: Concepts, Methodologies, Tools, and Applications, India, 2019, p. 22.

S. Mehla and S. Jain, "Rule Languages for the Semantic Web," in Emerging Technologies in Data Mining and Information Security, 2018.

L. Botha, T. Meyer, and R. Peñaloza, "A Bayesian Extension of the Description Logic $\mathcal{ALC}$." In European Conference on Logics in Artificial Intelligence, Cham, 2019.

S. Mehla and S. Jain, "Rule languages for the semantic web," in Emerging Technologies in Data Mining and Information Security, Singapore, 2019.

wikipedia, "RDFLib," [Online]. Available: https://en.wikipedia.org/wiki/RDFLib. [Accessed 10 Nov 2019].

N. Harrison and A. Paris, "Using Pattern-Based Architecture Reviews to Detect Quality Attribute Issues-An Exploratory Study," in Transactions on Pattern Languages of Programming III, Berlin, Heidelberg., 2013.