CONCEPT EXTRACTION FROM THE WEB OF THINGS KNOWLEDGE BASES

Mahda Noura
Technische Universität Chemnitz

Amelie Gyrard
Wright State University

Sebastian Heil
Technische Universität Chemnitz

Martin Gaedke
Technische Universität Chemnitz

CONCEPT EXTRACTION FROM THE WEB OF THINGS KNOWLEDGE BASES

ABSTRACT

Semantic web technologies are a major driver for semantic interoperability in IoT-generated data by using shared vocabularies in an ontology-driven approach. While there is a growing interest in standardization of ontologies for IoT, there is still a lack of common agreement for a specific IoT ontology. Numerous concepts and relations have been designed within existing ontologies to handle different features of IoT data. However, there are many redundant and overlapping concepts designed within existing standardizations and groups. We found that new ontologies constantly redesign the same concepts in IoT. Therefore, it is a challenge to unify these different IoT ontologies with redundant concepts. In this paper we investigate *what are the most used terms within IoT ontologies?* We identify the fourteen most popular ontologies within generic IoT and WoT domain. Analysis of popular concepts among these ontologies allows to automatically extract the information for ranking algorithms from information retrieval domain. This work will enable guiding ontology engineers to re-use existing ontologies and to unify ontologies, a required step to achieve semantic interoperability. Moreover, this work contributes towards building iot.schema.org.

KEYWORDS

Internet of things, Web of things, Semantic web of things, Semantic Interoperability, Ontology Engineering, Information retrieval, Knowledge Graph.

1. INTRODUCTION

The Internet of Things (IoT) has developed as an emerging technology in the recent years. IoT is defined as the connection of physical "things" and places using the Internet (Ashton & others 2009). With the increase in the number of IoT devices, the integration of IoT and the Web gradually started leading to the Web of Things (WoT) (Guinard et al. 2010). The WoT takes advantage of the universal accessibility of the Web. Data is generated by things and then exploited by web-based applications to monitor or control devices in different domains such as healthcare, home automation, transportation etc. According to a prediction provided by Cisco[1], the amount of annual global IP data traffic will reach 20.6 zettabytes by the end of 2021. It is noteworthy to point out that this data generated by the things are multi-modal containing different formats.

Interoperability in IoT is a key challenge due to the large heterogeneity of elements comprising IoT. According to (Noura et al. 2018), IoT interoperability should be handled in multiple levels. On the device level, challenges related to the processing capabilities of devices and different communication protocols must be addressed. The network level deals with providing a gateway to bridge between the different communication technologies. For syntactical level interoperability challenges related to the data format are dealt with. To handle the dynamic data models and schemas between heterogenous IoT devices and applications, the semantic level interoperability needs to be addressed. This can be realized by using Semantic Web of Things (SWoT) (Scioscia & Ruta 2009) paradigm to achieve an agreement on the meaning of the data by using shared vocabularies either in a schema form and/or in an ontology-driven approach (Gyrard et al. 2015).

Current works have proposed the use of unified ontologies as a solution to address the challenges related to interoperability of sensor data (Nambi et al. 2014; Gyrard et al. 2014). However, developing one comprehensive unified ontology for the IoT domain is challenging as there is a growing interest in standardization of ontologies to represent IoT devices and produced data. For example, W3C Semantic Networks

[1] https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.html

(SSN)[2] is the first initiative to address interoperability issues to describe sensor networks through an ontology. Sensor and devices are required to build WoT applications. The last release of the SSN ontology became a W3C recommendation in October 2017. It is a joint contribution with the Open Geospatial Consortium (OGC) standard, extending and improving the SSN ontology published in 2011. The W3C Web of Things (WoT) Interest Group[3] is designing a vocabulary to describe interactions between objects through the Web, a potential implementation is the WoT ontology. OneM2M[4], an international standard for Machine-to-Machine (M2M) designed the OneM2M ontology. Moreover, the ETSI Smart Appliances Reference (SAREF) ontology[5] was released initially in the domain of smart appliances for the smart home (e.g., smart refrigerators and ovens). However, it has been extended to cover other features of the IoT domain in general.

The existing schema.org vocabulary is planned to be extended for IoT devices and service. This extension is called iot.schema.org[6]. It has not been implemented yet, but collaboration is ongoing between different organizations. For developing the iot.schema.org a necessary step is to identify the most relevant concepts in a set of domain ontologies to unify multiple ontologies. However, many of the ontologies found in existing standardizations and different groups have many redundant concepts re-designed instead of reusing existing knowledge. Therefore, it is a challenge to unify these different IoT ontologies with redundant concepts, resulting in significant challenge for ontology experts and developers to re-use already designed ontologies, a required step to achieve semantic interoperability. As demonstrated by latest work such as (Seydoux et al. 2016; Agarwal et al. 2016) extracting the most popular concepts from an ontology catalogue is a manual task which results in poor ontology re-use by developers. An automated system can reduce the development time, increase re-use, and improve semantic interoperability among systems.

Therefore, to take a step towards facilitating ontology experts and developers to re-use existing ontologies and knowledge extraction experts to utilize the main concepts within IoT/WoT domain, we propose to automatize the task of detecting the most popular concepts within already designed IoT and WoT ontologies from a specific domain (e.g., smart home). We automatically extract the relevant knowledge from existing designed ontologies by using well-established techniques from information retrieval and machine learning domains. Importantly, in this work we are not aiming to enhance the state of the art in information retrieval, but rather to re-use existing state-of-the-art techniques 1) to find the most popular concepts and 2) the reused concepts in IoT and WoT domain. The outcome presented in this study will be beneficial in decreasing the time for discovering the most popular IoT concepts for re-use and simplifying the task of ontology experts to achieve semantic interoperability (Murdock 2016).

This article is structured as follows. Section 2 presents the related work, section 3 introduces the methodology for identifying the most popular concepts within IoT/WoT ontologies. The results of the study are deliberated in Section 4 and the discussion of the results are provided in Section 5. Finally, Section 6 concludes the paper and provides insights for future work.

2. RELATED WORK

Over the past few years, there have been some studies focusing on analysing IoT ontologies from different perspectives. For instance, the European research cluster on the IoT (IERC) released a set of best practices and recommendations for semantic interoperability (Things 2015). In (Ganzha et al. 2017), the authors survey the most popular ontologies used in (e/m) health and transportation or logistic domain. Gyrard et al. study the six different ontology-based validation tools (Parrot, WebVOWL, Oops, TripleChecker, LODE and Vapour) on 27 IoT ontologies with the aim to achieve semantic interoperability in IoT and WoT domain (Gyrard et al. 2018). This paper differentiates itself from the existing works in this category in that it analyses IoT ontologies to find the most popular concepts which have been used among a list of ontologies as well as the amount of concept re-use per ontology.

---

[2] https://www.w3.org/TR/vocab-ssn/

[3] https://www.w3.org/WoT/IG/

[4] http://www.onem2m.org/technical/onem2m-ontologies/

[5] https://sites.google.com/site/smartappliancesproject/ontologies/reference-ontology/

[6] https://iot.schema.org/

Information retrieval techniques have been used to find and rank non-IoT ontologies. For example, Swoogle[7], Watson[8], Sindince.com (Oren et al. 2008) and OntoKhoj (Patel et al. 2003) are search engines to help find relevant ontologies through user queries. They rank the resources in the ontologies using a link analysis and referrals between the ontologies, a method adapted from the PageRank algorithm. Nevertheless, this method of ranking is not suitable for IoT ontologies which are isolated and do not contain links from other ontologies since they would receive low ranks. Moreover, OntoKhoj and Sindice.com tools are no longer accessible and maintained. On the other hand, Falcons (Ontologies 2011) ranks concepts and ontologies using a popularity-based technique. However, this method focuses on ranking instances and does not focus on ranking ontology resources i.e., classes and properties. This tool is also not available anymore. AKTiveRank (Alani et al. 2006) ranks ontologies according to the ontology structure. It can retrieve a list of ontologies based on a user query search and then applies several analytic methods including Class Match Measure (CMM), Density Measure (DEM), Semantic Similarity Measure (SSM) and Betweenness Measure (BEW) to rate each ontology. General search services and algorithms have been developed for linked data applications, for instance, OntoSearch (Thomas et al. 2007) and OntoSelect. However, none of them are available anymore. Among the existing ontology search tools only Swoogle and Watson are still accessible on the Web, but they are not mature enough for identifying ontologies related to the IoT domain. This is mainly because of the differences between non-IoT ontologies and IoT ontologies. Most IoT ontologies are not even published online or do not follow semantic web practices for sharing and reusing the already designed domain knowledge. For instance, they lack labels or comments, domain or range, and properties for a given concept etc., Therefore existing ranking solutions and search engines are not suitable for IoT domain knowledge.

In the above-mentioned approaches IR algorithms are used to provide a ranked list of ontologies in response to a given user keyword which helps users to select the most appropriate ontology for re-use. Let us consider a scenario where a device provider wants to describe their new IoT device. There is no single ontology that is capable of describing the device (device capability, data, measurement, etc.) due to the isolated landscape of IoT ontologies. Therefore, the device provider must find the most popular concepts for each feature in different ontologies and aggregate them for its own purpose. This necessitates selecting the most popular concept among a set of ontologies rather than selecting the most popular ontology. In this regard, we find the most popular concepts and properties in a collection of IoT ontologies. Analysis of popular concepts among these ontologies allows to automatically extract the information as an input to the ranking algorithms from information retrieval domain. Our approach is similar to (Guha et al. 2015) which identifies the most popular types and relations in Schema.org.

Table 1. Qualitative comparison between existing information retrieval solutions.

Approach	**Input**	**Ranking criteria**	**Search terms**	**Tool availability**	**IoT ontology support**	**Output**
OntoKhoj, Sindice	keyword	semantic links, coverage	class, subclass, domain/range	✗	✗	ranked ontologies
Swoogle, Watson	keyword	semantic links, coverage	class, property, label, comment, literal	✓	✗	ranked ontologies
AKTiveRank	keyword	ontology structure	terms		✗	ranked ontologies
Falcons	keyword	concept popularity	class, property, labels	✗	✗	ranked ontologies
OntoSelect, OntoSearch2	keyword, corpus,	connectedness, ontology structure	class, label, property, ontology-title	✗	✗	ranked ontologies
Our approach	**ontologies**	**concept popularity**	**class, subclass, property, label, domain/range**	✓	✓	**popular concepts**

[7] http://swoogle.umbc.edu/2006/
[8] http://watson.kmi.open.ac.uk/WatsonWUI/

3. IDENTIFYING MOST POPULAR CONCEPTS IN IOT ONTOLOGIES

In this section we present the methodology that we employ for identifying the most popular concepts from a set of recent and relevant IoT ontologies. Figure 1 illustrates an overview of the workflow using BPMN. The popular concept identification architecture involves two roles: *Expert* are the actors in charge of conducting the popular concept identification and the *Analysis Toolchain* is a system role representing our tool chain for supporting popular concept identification through code scripts. An expert starts the process by using the Linked Open Vocabularies for Internet of Things (LOV4IoT) catalogue as input. Next, the analysis tool chain automatically identifies a list of most popular concepts as well as the frequency of the concept in all ontologies and the list of ontologies which have used the concept as output. In the following we describe the steps of this process in sequential order.

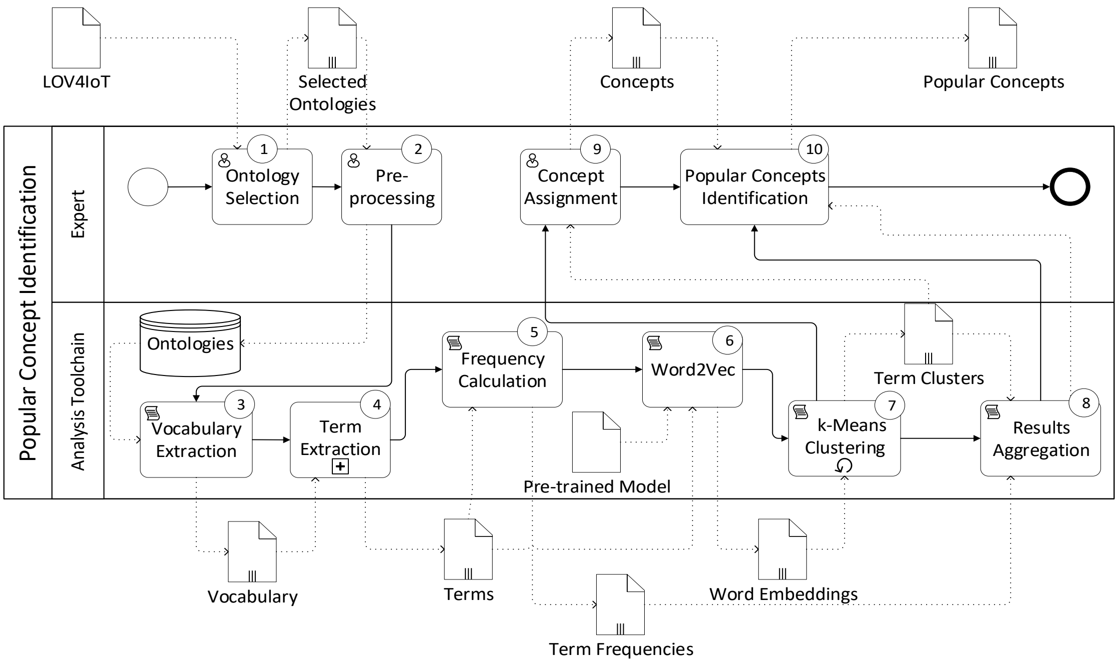

Figure 1. Process of Identification of Popular Concepts in IoT Ontologies

1. Ontology Selection: The LOV4IoT[9] catalogue is an input to the architecture which includes 391 ontology-based research projects representing different IoT domains such as home automation, smart cities, smart agriculture, healthcare, etc. Among these ontologies we selected the 14 most prominent ontologies (Table 2) related to the IoT and WoT domain. We focus on ontologies from standards with the following criteria's: the most cited, availability of the ontology code online, and the ability to work with a set of validation tools, taking (Gyrard et al. 2018) as the reference.

2. Preprocessing: A preprocessing step is required on the selected ontologies to improve the extraction results. The ontologies selected from the LOV4IoT catalogue are in various formats (RDF/XML, TTL, etc.). Herby, the preprocessing step takes care of converting the ontologies to the well-known TTL representation using Protégé. In the next step, the Unicode-encoded characters in the ontologies were converted to ASCII since several commonly used ontology parsers do not handle Unicode-encoded text correctly. Then, the preprocessed ontology resources are stored in a Virtuoso database by treating the ontologies as sets of triples.

3. Vocabulary extraction: Given a set of loaded ontologies in Virtuoso $O = \{o_1, o_2, \dots, o_n\}$ this step extracts the set of distinct *vocabulary* $V = \{w_1, w_2, \dots, w_v\}$. The ontology resources were parsed using OntoSpy[10] library in Python to identify all unique vocabularies from the ontology classes, subclasses, class properties, SKOS and labels defined by RDF.

[9] http://lov4iot.appspot.com/
[10] https://pypi.org/project/ontospy/

Table 2. Ontologies used in this work with name, number of triples and domain.

Name	#Concepts	Domain
BOnSAI	127	Smart building
IoT-O	60	Conceptual and functional requirements of IoT
OpenIoT	75	Concepts for IoT applications and testbeds
M3-lite	449	Unify sensor data
OneM2M	63	Representing IoT device and its functionality
Hachem	137	IoT heterogeneity and scalability
FIESTA-IoT	529	Unify existing IoT ontologies
WoT	35	Model and common representation for WoT
VICINITY	147	Model the WoT domain
SSN V1	119	Sensor resources and data collected through sensors
SSN V2	37	Sensor resources and data collected through sensors
SPITFIRE	122	Sensor, observation, and related concepts
VITAL	158	Sensor, measurement, time and locations concepts
Hypercat	15	Exposing information about IoT assets over the Web

4. Term Extraction: This step consists of two main phases. First the identified vocabulary V is split up on camel case or snake case into words called tokens, and perhaps certain characters such as punctuation marks are removed. Then filtering is performed to remove stop words. Stop words are the words that frequently appear in the text without having much content information (i.e., prepositions, conjunctions, etc.). The output is a list of identified terms $T = \{t_1, t_2, \ldots, t_T\}$.
5. Frequency Calculation: Given a set of terms $t \in T$, the frequency of the term in the set of ontologies O is calculated as $f_O(t)$. It represents the total number of occurrences of each unique term in all ontologies.
6. Word2vec: This step performs the training of the term embeddings and the process of building a *word2vec* model for all identified unique words. The *word2vec* algorithm is based on neural networks and builds a vocabulary from a pre-training text model and attaches the vector representations to each word. In addition, it calculates the cosine distance among each word. The genism python library was to implement *word2vec.* Around 20 of the terms were not part of the pre-trained model thus we removed those terms from the list of words. The output of this step is the word embedding vector space representation.
7. k-Means Clustering: By using the word embeddings from the previous step, we have the benefit to cluster similar words. Therefore, the main motivation of this step is to create a set of non-empty semantically coherent clusters from the original set of identified terms. For this purpose, the k-Means algorithm is used to perform clustering. The algorithm is an unsupervised machine learning algorithm which operates on the vector spaces. Given the value of K, the N terms are portioned into K distinct clusters based on the nearest mean calculation (Euclidean distance). The output consists of a text file with several clusters where each cluster contains a list of semantically coherent terms. Different values of K produce different number of clusters.
8. Results Aggregation: Using the term frequencies and term clusters as input this step calculates the total number of occurrences of the terms in different ontologies as well as the ontologies using this term per cluster.
9. Concept Assignment: In this step, given the term clusters the ontology experts manually assigned names to each cluster based on the semantic meaning of the cluster.
10. Popular Concept Identification: This step uses the concepts from the previous step and the results from step 8 to produce a text file as output with a list of ordered concepts based on the popularity as well as the list of ontologies that have used each concept.

4. RESULTS

In this section we report the results. Two sets of experiments were carried out on the 14 ontologies listed in Table 2: experiment *E1 identification of popular concepts* and experiment *E2 analysis of re-use*. In preparation of E1 and E2, the set of IoT ontologies were loaded in a triple store comprising approximately 2k concepts. We carried out the experiments using a set of python scripts.
For experiment E1, the identification of popular concepts, we extracted the 20 most popular concepts utilized among IoT ontologies as described by the process illustrated in 3.

The vocabulary extraction in step 3 yielded 2073 unique vocabulary items, representing the names of classes, properties and SKOS-concepts in the ontologies. Subsequent term extraction in step 4 reduced this numbers to 958, because the splitting of camel-case and snake-case splits unique compound identifiers turning them into more frequent general terms. The average number of terms per ontology was calculated at 148. Step 5, the frequency calculation, produced the number of occurrences of these terms in all ontologies. The most frequent term was *sensor* with 183 occurrences. The Word2vec algorithm in step 6 employed the pre-trained Google News Model which is derived from a large corpus of words from Google News. This text corpus contains about 3 million words from 20 different news groups represented by 300-dimensional vectors using negative sampling. We then performed clustering experiments using the k-Means algorithm with different values of k to achieve a set of semantically consistent clusters. The value of k=55 was found to produce the most meaningful and semantically consistent clusters with regards to the IoT domain. For k=55, the average size of each cluster is around 16. For each cluster, we assigned a concept that best describes the cluster members and their semantic relationship in step 9.
Table 3 shows the results of these 20 most popular concepts used in the IoT domain. The concepts are listed by aggregated term frequencies of all terms of the corresponding cluster among the different ontologies. Please note that the example terms for each concept are shown to provide an idea of the cluster terms, however, they are not the complete set of all terms in that cluster.

Table 3. The 20 most popular IoT concepts and term frequency among 14 different IoT ontologies.

Concept	**Term frequency in different ontologies**
Unit of measure (*inch, centimeter, millimeter, tone, kilometer*)	**268** terms found in **11** ontologies
Sensor (*monitoring, detection, reactive, tagging, filtering*)	**238** terms found in **11** ontologies
Chemical (*nitrogen, co2, monoxide, ion, carbon, radiation*)	**215** terms found in **4** ontologies
Physical parameters (*luminosity, radiance, sensitivity, selectivity*)	**192** terms found in **10** ontologies
System (*platform, device, software, computing, network*)	**173** terms found in **12** ontologies
Mathematical Terms (*limit, count, level, per, threshold, min, max*)	**171** terms found in **11** ontologies
Environment Parameters (*humidity, temperature, wetness, moisture*)	**170** terms found in **6** ontologies
Abbreviation (*iot, conn, gps, lat, var, uri, lo, tv, id, ecg, pc, attr, io*)	**143** terms found in **11** ontologies
Energy (*solar, renewable, hvac, electric, fuel, battery, charger*)	**140** terms found in **10** ontologies
Geometry (*circle, angular, surface, pattern, angle, line, radius, length*)	**136** terms found in **9** ontologies
Metadata (*impact, participant, contributor, topic, result, role, creator*)	**132** terms found in **11** ontologies
Measurement (*observation, transduce, resolution, accuracy, precision*)	**125** terms found in **11** ontologies
Mathematical Estimation (*probability, forecasting, equivalent, relative*)	**116** terms found in **10** ontologies
Service (*services, action, work, quality, care*)	**107** terms found in **11** ontologies
Actuator (*actuation, actuate, throttle, motion, acceleration, mechanical*)	**106** terms found in **8** ontologies
Controlling (*operating, controlled*)	**98** terms found in **11** ontologies
Environment (*ecosystem, climate, environmental, region, area, state*)	**84** terms found in **9** ontologies
Time (*starting, current, day, year, last, end, initial, date, schedule*)	**80** terms found in **9** ontologies
Location (*square, indoor, parking, space, building, grounds, recreation*)	**76** terms found in **10** ontologies
Status (*condition, situation, priority, term, returns*)	**68** terms found in **11** ontologies

In experiment E2, the analysis of re-use, the ontologies were analyzed to identify the extent of re-use of concepts among them. We considered concept re-use in OWL ontologyies by 3 mechanisms: 1) direct re-use through the use of classes, properties or skos-concepts from other ontologies (namespaces), 2) extension of classes or properties from other ontologies through sub-classing (using `rdfs:subClassOf`) and, 3) exact reuse of classes and properties from other ontologies (`owl:sameAs, owl:EquivalentClass, owl:equivalentProperty`)

To identify these 3 cases of re-use, the following process was employed:
First, we identified the base URI of each ontology from the subject of the triple (subject, rdf:type, owl:Ontology). Next, we extracted the subject-URIs of all classes, properties, skos-concepts and object URIs of all subClassOf-triples for classes and properties. The URIs from this combined list were then matched against the base URI of the ontology. If the URIs did not match, the respective concept was from a different namespace than the ontology itself and therefore re-used from another ontology. By comparison of the number of overall concepts in the ontology with the number of concepts identified as re-used from other ontologies, we calculated the percentage of re-use in each ontology. Our findings are reported in Table 4. In addition, we

indicate the source of re-use in terms of ontologies from which concepts were re-used as well as the number of concepts taken from that source per ontology in our experimental dataset.

Table 4. The extent of re-use for each IoT ontology from other ontologies.

Ontology	# of Concepts Re-used	Total Concepts	% Re-use	Source of re-use
Fiesta-IoT	462	529	87	M3-lite:359,SSN:31,IoT-lite:63, OneM2M:9
M3-lite	245	449	54.56	SSN, IoT-lite
Vital	89	158	56.32	MSM[11]:49, SSN:31
Vicinity	74	147	50.34	SSN:53, SAREF:3, WoT:1
SSNv2	5	37	13.51	SOSA:3
IoT-O	11	60	18.33	SAN:1
WoT	2	35	5.71	-
openIoT	3	75	4.00	SSN:3
BOnSAI	3	127	2.36	CoDAMoS:3
SPITFIRE	1	122	1.63	SSN:2
SSNv1	0	119	0	-
oneM2M	0	63	0	-
Hachem	0	137	0	-
Hypercat	0	15	0	-

Presently, 10 out of the 13 ontologies (76%) re-use as a minimum one term from another ontology and 5 ontologies (38%) have at least one of their terms re-used. The SSNv1 is the most commonly re-used ontology; 5 different ontologies re-used concepts from it. The FIESTA-IOT ontology has 100% re-use from 4 different ontologies. Notably, 6 of the ontologies do not re-use any IoT ontologies at all.

DISCUSSION

The experimental results discussed in the previous section illustrates the most relevant concepts and the extent of concept reuse among a set of IoT ontologies. E1 indicates that the *unit of measurement* is the most frequent concept used within these ontologies. The reason could simply be that there exist many different measuring units. Also, units are essential to describe IoT data sent from the sensors and actuators, describing effects, preconditions, input and output levels etc. Among the set of ontologies, the M3-lite focuses on different types of *unit of measurements* (*Unit*) and alone includes 56 units. Moreover, according to Table 3 there are some rather specific concepts like *chemical, environmental parameter, geometry*, and *Actuator* as well as many generally used concepts which are used in at least 9 different ontologies. The popularity of the concept depends mostly on the ontology domain, obviously a health ontology will yield more health-related concepts. Since the ontologies that we have used in this study are generic IoT/WoT ontologies, the generic concepts have a higher popularity. Moreover, it can be noted that the *Sensor* concept is more popular than *Actuator*. The reason may lie in the fact that in the IoT market there are more sensors compared to actuators.

In experiment E2, the analysis of re-use, the results can be divided into two groups: above 50 % (very high), and less than 18%. There are 4 main ontologies in the first group: FIESTA-IOT, M3-lite, Vital and Vicinity and then there is a very large gap, then the second group. We believe the extent of reuse in the first category is high because they were designed with the aim to unify the existing ontologies to achieve semantic interoperability. For instance, the FIESTA-IOT ontology has 87% re-use from other ontologies. This is because it was created with the target to re-use concepts from IoT-lite, M3-lite, SSN, oneM2M. Similarly, M3-lite and

[11] https://lov.linkeddata.es/dataset/lov/vocabs/msm/

Vicinity was created with the target to re-use and aggregate existing ontologies whenever possible to achieve semantic interoperability. VITAL is also an extension of the SSNv1 ontology. However, the other ontologies have a very low reuse, even four of the ontologies do not reuse any concepts.
Among the four ontologies that do not reuse concepts, SSNv1 is the most commonly re-used ontology among all the analyzed ontologies. This is due to the popularity of this ontology among different IoT initiatives. SSNv1 can be considered as a de-facto standard ontology. The most re-used concept from SSNv1 is the *Sensing Device,* which is a sensor that reports measurements and observations of real world phenomena. The identification of popular concepts, could obviously be integrated within iot.schema.org.

5. CONCLUSION

This work presents an analysis on IoT/WoT ontologies which is first of its kind. The main objective of this research was mainly to find the most relevant concepts and the extent of concept reuse from a set of IoT/WoT generic ontologies to achieve semantic interoperability with a focus to improve the reuse of most popular concepts in this domain. The evaluation results illustrate that unit of measure is the most popular IoT concept and 71% of the studied ontologies have less than 18% concept reuse and 20% have no concept reuse.
The main target groups of this research are ontology experts, developers, researchers willing to discover, and study already designed ontologies within the IoT domain as well as knowledge extraction experts who want to utilize the main concepts within the IoT domain. The outcome presented in this study can be beneficial in decreasing the time for discovering the most popular IoT concepts for re-use and simplifying the task of ontology experts to achieve semantic interoperability. The benefit of the employed methodology is that it is generic enough to be applied to any domain. Moreover, the application of this work can assist the development of iot.schema.org. While conducting this study we noticed that numerous concepts overlap within the ontologies. Therefore, in the future, we plan to extend the analysis by considering identifying the redundant concepts designed within IoT and WoT ontologies.

REFERENCES

Agarwal, R. et al., 2016. Unified IoT ontology to enable interoperability and federation of testbeds. In *Internet of Things (WF-IoT), 2016 IEEE 3rd World Forum on*. pp. 70–75.

Alani, H., Brewster, C. & Shadbolt, N., 2006. Ranking ontologies with AKTiveRank. In *International. Semantic Web Conference*. pp. 1–15.

Ashton, K. & others, 2009. The internet of things. *RFID journal*, 22(7), pp.97–114.

Ganzha, M. et al., 2017. Semantic interoperability in the Internet of Things: An overview from the INTER-IoT perspective. *Journal of Network and Computer Applications*, 81(October 2017), pp.111–124.

Guha, R. V., Brickley, D. & MacBeth, S., 2015. Schema.org: Evolution of Structured Data on the Web. *Queue*, 13(9), pp.10–37. Available at: http://dl.acm.org/citation.cfm?doid=2857274.2857276.

Guinard, D., Trifa, V. & Wilde, E., 2010. A resource oriented architecture for the web of things. In *Internet of Things (IOT)*. pp. 1–8.

Gyrard, A., Bonnet, C. & Boudaoud, K., 2014. Enrich machine-to-machine data with semantic web technologies for cross-domain applications. *2014 IEEE World Forum on Internet of Things, WF-IoT 2014*, pp.559–564.

Gyrard, A., Datta, S.K. & Bonnet, C., 2018. A Survey and Analysis of Ontology-Based Software Tools for Semantic Interoperability in IoT and WoT Landscapes. , pp.86–91.

Gyrard, A., Serrano, M. & Atemezing, G.A., 2015. Semantic web methodologies, best practices and ontology engineering applied to Internet of Things. In *Internet of Things (WF-IoT), 2015 IEEE 2nd World Forum on*. pp. 412–417.

Murdock, P., 2016. Semantic Interoperability for the Web of Things. *ResearchGate*, (October), pp.1–19.

Nambi, S.N.A.U. et al., 2014. A unified semantic knowledge base for IoT. *WF-IoT 2014*, pp.575–580.

Noura, M., Atiquzzaman, M. & Gaedke, M., 2018. Interoperability in Internet of Things: Taxonomies and Open Challenges. *Mobile Networks and Applications*, pp.1–14.

Ontologies, W., 2011. Correspondence Falcons Concept Search : A Practical Search Engine for. , 41(4), pp.810–816.

Oren, E. et al., 2008. Sindice . com : A Document-oriented Lookup Index for Open Linked Data. *Int.J.Metadata Semant.*

Ontologies, 3(11).

Patel, C. et al., 2003. OntoKhoj: A Semantic Web Portal for Ontology Searching, Ranking, and Classification. *WIDM 2003*, pp.58–61.

Scioscia, F. & Ruta, M., 2009. Building a Semantic Web of Things: Issues and perspectives in information compression. *ICSC 2009 - 2009 IEEE International Conference on Semantic Computing*, pp.589–594.

Seydoux, N. et al., 2016. IoT-O, a Core-Domain IoT Ontology to Represent Connected Devices Networks. , pp.561–576.

Things, E.R.C. on the I. of, 2015. Internet of Things - IoT Semantic Interoperability: research challeges, best practices, recommendations and next steps. , 1, p.48.

Thomas, E., Pan, J.Z. & Sleeman, D., 2007. ONTOSEARCH2: Searching ontologies semantically. *CEUR Workshop Proceedings*, 258.