2017 IEEE 37th International Conference on Distributed Computing Systems

Towards a RISC Framework for Efficient Contextualisation in the IoT

Towards a Unified Contextualisation Framework

Dimitrios Georgakopoulos [†], Ali Yavari [*†], Prem Prakash Jayaraman [†], Rajiv Ranjan [×]
[*] RMIT University, Melbourne, Australia
[†] Swinburne University of Technology, Melbourne, Australia
[×] Newcastle University, Newcastle, United Kingdom
dgeorgakopoulos@swin.edu.au, mail@aliyavari.com, pjayaraman@swin.edu.au, rranjans@gmail.com

Abstract — **The Internet of Things (IoT) is a new internet evolution that involves connecting billions of internet-connected devices that we refer to as IoT things. These devices can communicate directly and intelligently over the Internet, and generate a massive amount of data that needs to be consumed by a variety of IoT applications. This paper focuses on the automatic contextualisation of IoT data, which also involves distilling information and knowledge from the IoT aiming to simplify answering the following fundamental questions that often arises in IoT applications: Which data collected by IoT are relevant to myself and the IoT Things I care for? Related work around context management and contextualisation ranges from database techniques that involve query re-writing, to semantic web and rule-based context management approaches, to machine learning and data science-based solutions in mobile and ambient computing. All such existing approaches have two main aspects in common: They are highly incompatible and horribly inefficient from a scalability and performance perspective. In this paper, we discuss a new RISC Contextualisation Framework (RCF) we have developed, implemented key aspects of, and assess its scalability. RCF provides fundamental contextualisation concepts that can be mapped to all existing contextualisation approaches for IoT data (and in this sense, it provides a common denominator that unifies the contextualisation space). RCF can be easily implemented as a cloud-based service, and provides better scalability and performance that any of the existing content management and contextualisation approaches in the IoT space.**

Keywords—Context; Contextualisation; Scalability; Internet of Things

I. Introduction

The Internet of Things (IoT) currently incorporates approximately 15B IoT devices, and there are estimates of 50B+ by 2020 [1], [2, p. 4]. The aim of the IoT is to utilise such devices, which produce data and are capable of actuating and affecting the physical world. IoT devices produce a tremendous amount of data, which we refer to as IoT data. IoT data underpins the development of IoT applications that support many novel products and services aiming to make Cities, Health, Manufacturing (Industry 4.0), Energy Grids, and Agriculture smarter.

Many IoT applications typically follow the Observation, Orientation, Decision and Action (OODA) loop paradigm [3]. To explain this further, consider an IoT-based application for a smart city that, among other services, finds and recommends parking spaces to drivers who commute to work. At the Observation phase this application collects a variety of IoT data from a variety of sources, including (1) traffic, parking, roadside, and public transport sensors, (2) scheduling and route information from public transport databases, and (3) location and navigation information for its users (i.e. the drivers, vehicles, buses, trams, and trains). Orientation for the parking recommendation service involves contextualising all such information to recommend a specific parking spot to each driver/car. Contextualisation involves filtering out all the IoT data that do not relate to parking recommendation. Orientation for the parking recommendation service next aggregates the remaining IoT parking-related information with the destination selected by each driver. Finally, it correlates the preferences of the driver (e.g. parking cost and distance from destination) to a recommendation. The driver then decides on a specific recommendation, and sets the selected parking space in his navigation system as the destination. Many other IoT applications and related services they provide have OODA loop data processing structures.

One of the important challenge towards delivering virtually all IoT services and products is contextualising IoT data, i.e. deducing information and knowledge that is relevant to a specific person and the IoT Things that he/she identifies (i.e. cares for), doing this for groups of people and IoT Things (e.g. those that belong to a specific organisation or social network), and doing this for dynamic groups of these (e.g. when a vehicle drives down the road relevant road-side sensors change, roads change and drivers change). Therefore, the completeness and efficiency of large scale contextualisation is highly important for the success of IoT.

There are many existing research solutions for context management and contextualisation that range from database techniques that involve query re-writing, to semantic web and rule-based context management approaches, to machine learning and data science-based solutions in mobile and ambient computing. Virtually all existing contextualisation

1063-6927/17 $31.00 © 2017 IEEE
DOI 10.1109/ICDCS.2017.308

approaches are incompatible as they consider different notions of content and advocate incompatible contextualisation techniques. Furthermore, as none of these existing approaches has been designed to support the scale of IoT data, they are inefficient from a scalability and performance perspective.

In this paper we discuss the ConTaaS framework, the mapping of existing contextualisation approaches to this framework, and illustrate the scalability it provides using sample IoT large-scale applications, such as Smart City [4]) referred as Internet-scale data.

II. CONTEXTUALISATION FRAMEWORK

Context and Contextualisation in IoT and Internet-scale data have been defined as follows [3], [5]: *Context or contextual information is any information about any entity that can be used to effectively reduce the amount of reasoning required (via filtering, aggregation, and inference) for decision making within the scope of a specific application.*

Contextualisation is the process of identifying the data relevant to an entity based on the entity's contextual information. In other words, contextualisation in an IoT application is the techniques that exclude irrelevant IoT data (and in some cases, also includes relevant IoT data) from consideration in processing and dissemination. Therefore, contextualisation can increase the relevance of information considered by an IoT application and significantly reduce the volume, velocity, and variety of the data considered and processed by the IoT application

Fig. 1. depicts a RISC *Contextualisation Framework* (RCF) that is being designed and developed for contextualising IoT data. In RCF, data coming from IoT sensors and other devices are Raw Data as depicted in Fig. 1.. RCF employs a data triple format for processing and storing all IoT data. IoT data triples can be mapped to virtually *any* knowledge representation model. Fig. 1. depicts RDF [6] as the knowledge representation model of choice for contextualisation. However, triples can also be mapped to database schemas, OWL ontologies, rules, as well as other knowledge representation models with a corresponding degree of expressive power.

Raw Data triplets are used for contextualisation that involves applying contextualisation operations. There are three types of contextualisation operations in RCF:

- *Contextual Filter*: This filters the IoT data triplets based on the current context. For instance, when a driver is looking for parking in Hawthorn (a Melbourne suburb), the contextual filter excludes from further processing and contextualisation output all IoT data received from a parking sensors located in other suburbs, other Australian cities and other countries.
- *Contextual Aggregation*: This aggregates the IoT data based on the contextual similarities and relevance. For example, if all drivers searching for parking spots in Melbourne have SUV vehicles then we can aggregate SUV and Melbourne data and treat their combination as a new context.
- *Contextual Interference*: This is a more complex operation and is based on reasoning on the data and their relations. For example, recommending a parking space to a senior citizen may involve using the driver's age to compute walking distance to the selected parking spot.

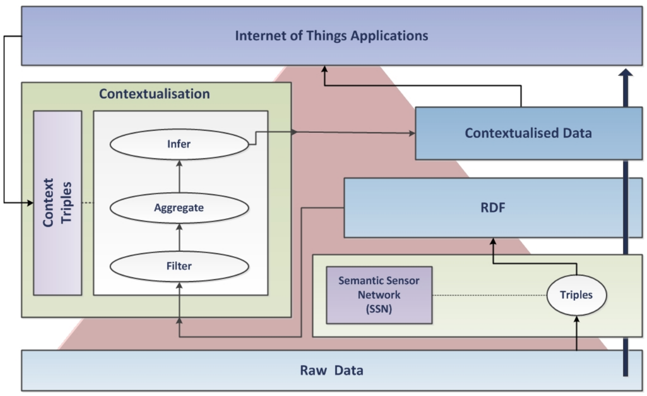

Fig. 1. RCF

ConTaaS [3] is a RCF implementation developed in Java and deployed on Amazon EC2 [7]. ConTaaS disseminates the filtered, aggregated, and inferred data (contextualised) to IoT applications using ARQ [8] Query Engine for SPARQL [9].

III. UNIFYING THE CONTEXTUALISATION SPACE

In Section II we discussed the way in which RCF can accommodate virtually any knowledge representation model for IoT data. RCF can achieve a high degree of interoperability via the use of an ontology that formally describes the IoT data processed by each IoT application. For example, RCF can use the Semantic Sensor Network (SSN) ontology [17] that has been designed for IoT applications. RCF is also compatible with less expressive knowledge representation approaches such as schemas in relational databases and RDF. Therefore, RCF is designed to and can employ a variety of knowledge representation models ranging from RDF, to Semantic Web, to databases and rules.

Similarly, many data processing and analysis techniques developed for contextualisation based on clustering, regression, sentiment, statistics, and fuzzy approaches [10]–[18] can be mapped to the Contextual Filtering, Aggregation, and Inference operations provided by RCF. While RCF provides for unification of knowledge representation and contextualisation techniques, to be a truly RISC framework it must also provide exceptional performance and scalability.

IV. CONTEXTUALISATION SCALABILITY

We use the term Internet-scale data to refer to the massive data generated from IoT devices and services (e.g. sensors, and smart phones). Contextualisation can increase the performance and efficiency of Internet-scale data processed by IoT applications. Although contextualisation can be viewed as a subclass of data analytics, many of the latest high-performance processing techniques for Big Data, such as MapReduce [4], are not ideal for such IoT contextualisation because they fall short in supporting the incremental data processing [5] requirements of contextualisation, and the near real-time requirements of many IoT applications.

RCF's Contextual Filter, Aggregation, and Inference operation can drastically reduce the amount of data they produce based on the relevance of the IoT data to the user at hand and the IoT Things each user cares for.

V. Sample Applications

In this section, we present IoT applications in three different domains that use RCF to meet the near real-time requirement needs of IoT.

Current applications consider context, however they cannot cope with Internet-scale data. For example, with smart cities there are more and more sensors appearing such as parking sensors, connected cars etc. The ability to capture data from all these sensors, analyse them and provide recommendations in near real-time is a complex problem. Furthermore, this problem gets more complicated every day with the addition of new sensor data sources.

By contextualising IoT data, the system can provide a contextualised service that takes into consideration multiple contexts originating from the user, the car and the Smart City (shopping centre space), by continuously computing relevant context, providing instant response to common parking queries and matching shared context among multiple users to efficiently answer queries. For example, consider the smart parking scenario. By incorporating the user, the car and the Smart City (shopping centre space) context and by continuously computing relevant context using Contextualisation, the Smart Parking application can provide real-time response to common parking queries. Using contextualisation, shared context among multiple entities (e.g. users) can be matched to efficiently respond to queries.

Using the RFC framework, for the parking scenario, we are able to perform contextualisation over IoT data enabling the application to respond to queries in real-time. The RFC group-related context shared between entities (including users, cars, and city) in order to support real-time resolution of parking recommendation queries. In [3], we have experimentally demonstrated that IoT applications such as parking recommendation can cope with Internet scale data using RFC framework. Contextualisation has the potential to improve data and query processing in Internet scaled IoT applications such as Smart Parking [Fig. 2.] by consideration of the data relevant to the contexts.

A. *Smart Farming*

Improving farm productivity requires crop performance to be understood and forecast under a wide variety of environmental, soil, fertilisation, and irrigation conditions. Productivity of a farm can be enhanced by determining which crop variety has produced the greatest yield under similar soil, climate, fertilisation, and irrigation conditions. The same data-driven approach to crop selection can also address climate change, resource constraints (water, labour, and energy shortages), and societal concerns around issues such as animal welfare, fertilizers, and environment that often impact agricultural production. ConTaaS has the potential to reduce the complexity of the data analysis query and extract valuable knowledge from the sensory data coming from the fields [19].

With the advent of smart farming, we are generating more and more data that is valuable to understand crop performance. However, timely processing of this data and responding to any events is a critical step. For example, detection of a frost event needs to be instantaneous in order to provide adequate response to protect the crop from frost attack. Any delay in detection or response can result in crop damage and potentially heavy losses.

Contextualisation in this application can help distil valuable information relevant to the crop (IoT data to the user at hand and the IoT Things each user cares for) such as the location, the farmer and the type of crop which will in turn support timely detection and response.

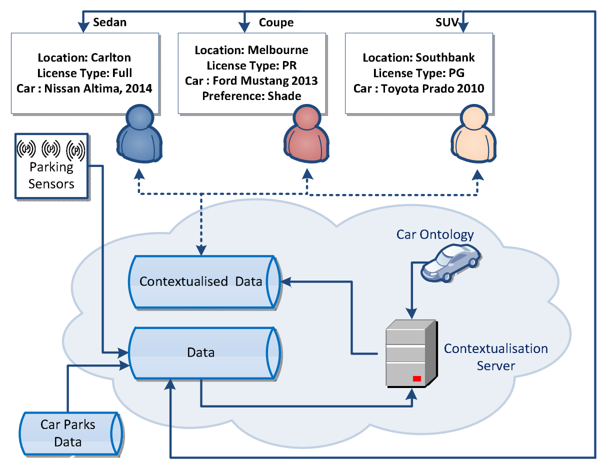

Fig. 2. Smart Parking

B. *Smart Healthcare*

Consider an outbreak of Ebola virus disease that originated in March 2014 from the west of Africa. After five months and more than 4500 death reports, the World Health Organization declared this outbreak an international public health emergency. To stop Ebola virus transmission that occurs via physical contact, it is necessary to do the following: 1) diagnose the virus as soon as possible, 2) isolate the patients by limiting contact with other people, and finally 3) start infection control and treatment. Ebola virus disease's most common symptoms are fever, fatigue, loss of appetite, vomiting, diarrhoea and headache. Speeding up diagnosis by identifying any person who has all or most of these symptoms and determining whether this person has been travelling in a high-risk area during a time interval could be potentially lifesaving. With current advances in mobile smart phone and wearable technology, we assume the possibility of collecting data from people including their location (with due concerns to privacy and security [20], [21]). Additionally, assume that we have the records of the symptoms of people that are manually entered into a database via health applications, collected from hospitals and during medical checks, and/or from sensors and wearable devices such as smart watches [22]. Such data from citizens will be massive. Furthermore, for any application we may need to frequently repeat the data analysis process on the entire dataset. Managing a dataset of this kind is resource demanding and its analysis requires sophisticated computing resources.

ConTaaS has the potential to solve this problem by reducing the complexity of the data analysis query and extract valuable knowledge from such data.

VI. Future Research Direction And Conclusion

In the current design of ConTaaS implementation, the context information needs to be added manually. However, it is often necessary to expand contexts with new information and reduce context information that is no longer relevant. Moreover, the smart things in the IoT ecosystem are also capable of generating new context as they evolve over time. For example, dynamic context change is common in the presence of mobility, change of interest or task at hand, or the change of the environment context entities are subject to. A future direction of RCF involves the development, mapping, and integration of additional machine learning algorithms for contextualisation to allow for automated or semi-automated context information collection and processing.

Another area of work very relevant to RCF is to investigate context verification techniques. The current assumption in RCF is that the context provided (manual or dynamic) is accurate. However, this is not always the case especially in an IoT ecosystem. Hence, new techniques need to be developed that can differentiate good and bad contexts.

On the positive side, RCF and its ConTaaS implementation have produced performance and scalability results that indicate that the proposed RCF framework and related contextualisation techniques scale linearly with large datasets. The framework described in this paper is designed to be incremental, but we did not address incremental functionalities of the algorithms. This incremental nature of RCF can help to investigate more complex reasoning by connecting other domain specific ontologies to SSN ontology. RCF is a significant development in context management and contextualisation and provides a solid foundation for further ground breaking research.

References

[1]. J. Rivera and R. van der Meulen, "Gartner Says The Internet Of Things Installed Base will grow to 26 Billion Units By 2020," Stamford Conn Dec., vol. 12, 2013.

[2]. E. Akhavan-Rezai, M. F. Shaaban, E. El-Saadany, and F. Karray, "Online intelligent demand management of plug-in electric vehicles in future smart parking lots," IEEE Syst. J., vol. 10, no. 2, pp. 483–494, 2016.

[3]. A. Yavari, P. P. Jayaraman, D. Georgakopoulos, and S. Nepal, "ConTaaS: An Approach to Internet-Scale Contextualisation for Developing Efficient Internet of Things Applications," in Proceedings of the 50th Hawaii International Conference on System Sciences, 2017.

[4]. P. Neirotti, A. De Marco, A. C. Cagliano, G. Mangano, and F. Scorrano, "Current trends in Smart City initiatives: Some stylised facts," Cities, vol. 38, pp. 25–36, 2014.

[5]. A. Yavari, P. P. Jayaraman, and D. Georgakopoulos, "Contextualised service delivery in the Internet of Things: Parking recommender for smart cities," in 3rd IEEE World Forum on Internet of Things, WF-IoT 2016, Reston, VA, USA, December 12-14, 2016, 2016, pp. 454–459.

[6]. G. Klyne and J. J. Carroll, "Resource description framework (RDF): Concepts and abstract syntax," 2006.

[7]. E. Amazon, "Amazon elastic compute cloud (Amazon EC2)," Amaz. Elastic Compute Cloud Amaz. EC2, 2010.

[8]. A. Seaborne, "ARQ-A SPARQL Processor for Jena," Obtained Internet Httpjena Sourceforge NetARQ, 2010.

[9]. E. Prud'Hommeaux, A. Seaborne, and others, "SPARQL query language for RDF," W3C Recomm., vol. 15, 2008.

[10]. C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Context aware computing for the internet of things: A survey," Commun. Surv. Tutor. IEEE, vol. 16, no. 1, pp. 414–454, 2014.

[11]. C. Baladron, J. M. Aguiar, B. Carro, L. Calavia, A. Cadenas, and A. Sanchez-Esguevillas, "Framework for intelligent service adaptation to user's context in next generation networks," Commun. Mag. IEEE, vol. 50, no. 3, pp. 18–25, 2012.

[12]. P. Castro and R. Munz, "Managing context data for smart spaces," IEEE Pers. Commun., vol. 7, no. 5, pp. 44–46, 2000.

[13]. M. Baldauf, S. Dustdar, and F. Rosenberg, "A survey on context-aware systems," Int. J. Ad Hoc Ubiquitous Comput., vol. 2, no. 4, pp. 263–277, 2007.

[14]. H. Cao et al., "Context-aware query suggestion by mining click-through and session data," in Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 2008, pp. 875–883.

[15]. L. Feng, P. M. Apers, and W. Jonker, "Towards context-aware data management for ambient intelligence," in International conference on database and expert systems applications, 2004, pp. 422–431.

[16]. C. K. Hess and R. H. Campbell, "A context-aware data management system for ubiquitous computing applications," in Distributed Computing Systems, 2003. Proceedings. 23rd International Conference on, 2003, pp. 294–301.

[17]. M. Amoretti, S. Copelli, F. Wientapper, F. Furfari, S. Lenzi, and S. Chessa, "Sensor data fusion for activity monitoring in the PERSONA ambient assisted living project," J. Ambient Intell. Humaniz. Comput., vol. 4, no. 1, pp. 67–84, 2013.

[18]. P. Vajirkar, S. Singh, and Y. Lee, "Context-aware data mining framework for wireless medical application," in International Conference on Database and Expert Systems Applications, 2003, pp. 381–391.

[19]. P. P. Jayaraman, A. Yavari, D. Georgakopoulos, A. Morshed, and A. B. Zaslavsky, "Internet of Things Platform for Smart Farming: Experiences and Lessons Learnt," Sensors, vol. 16, no. 11, p. 1884, 2016.

[20]. A. Yavari, A. Soltani Panah, D. Georgakopoulos, P. Prakash Jayaraman, and R. van Schyndel, "Scalable Role-based Data Disclosure Control for the Internet of Things," in Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on, 2017.

[21]. P. P. Jayaraman, X. Yang, A. Yavari, D. Georgakopoulos, and X. Yi, "Privacy preserving Internet of Things: From privacy techniques to a blueprint architecture and efficient implementation," Future Gener. Comput. Syst., p. , 2017.

[22]. T. Shah, A. Yavari, S. S. Karan Mitra, P. P. Jayaraman, F. Rabhi, and R. Ranjan, "Remote health care cyber-physical system: quality of service (QoS) challenges and opportunities," IET Cyber-Phys. Syst. Theory Appl., vol. 1, no. 1, pp. 40–48, 2016.