Detecting situations of importance with Stream Reasoning on live health IoT data

Mathieu Bourgais [a,*], Franco Giustozzi [a] and Laurent Vercouter [a]
[a] *Normandie Univ, INSA Rouen Normandie, LITIS, 76000 Rouen, France*
E-mail: mathieu.bourgais@insa-rouen.fr

Abstract. The development of Internet of Things (IoT) creates large amount of data which may be used by decision making systems in a variety of domains. In particular, in the field of health monitoring, it enables to follow the medical state of a patient hospitalized at home in real-time. An important challenge is to represent and interpret these data with a high-level model in order to have a better understanding of the overall medical state of a patient, taking into account the context of these data. This article overcomes this challenge by using Stream Reasoning techniques associated to an ontological representation of the medical context of a patient to understand her situation. This permits to combine in real time static knowledge stored in an ontology and dynamic information provided by smart sensors. To facilitate this process, constraints and situations concepts are introduced to ease the translation of expert knowledge into logical queries. The paper concludes with a discussion on the coverage of the proposed ontology and an experimental analysis of real body temperature data to illustrate how situations may be detected.

Keywords: Stream Reasoning, IoT, Medical data, Ontology

1. Introduction

The increasing development of Internet of Things (IoT) [1] provides large amount of data collected by smart sensors in numerous contexts : Industry 4.0 [2], Smart Cities [3], or medical monitoring [4]. The challenge with these data is to use them with efficiency to feed decision making systems, especially making the difference between background noise and valuable information. In the medical field, this means taking into account the patient's context in real time in order to help physicians with a decision making system that could make the difference between a real state of danger and a fake alert. Moreover, as these sensors generate data in real time at a high frequency, an important problem is to raise the right amount of alerts, neither too many (raising alerts when there are actually no problems) nor too few (missing a relevant sign indicating a patient's critical condition).

In order to raise an alert, data obtained from patients must be processed and integrated. This is a challenging task as the data is highly heterogeneous, coming from multiple type of sensors, with various temporal resolutions and different underlying meanings. Semantic Web technologies have proved their efficiency to deal with these issues [5]. Semantic Web is an extension of the World Wide Web which combines knowledge engineering and AI methods to represent, integrate, and reason upon data and knowledge through ontologies and rules. In computer science, an ontology is considered as "an explicit specification of a conceptualization for a domain of interest" [6]. Ontologies emerge as a pertinent method to represent knowledge of any kind, such as the medical domain, in a machine-interpretable way. Furthermore, reasoning on ontologies allows to transform raw observations collected by sensors into higher-level abstractions, such as situations of interest. These situations are meaningful for humans

*Corresponding author. E-mail: mathieu.bourgais@insa-rouen.fr.

1570-0844/$35.00 © 0 – IOS Press and the authors. All rights reserved

and provide a better understanding about the physical world, helping to support decision making tasks. However, current solutions for reasoning on ontologies were traditionally developed for static or slow changing data which does not correspond to the highly dynamic nature of data in the medical domain [7]. To tackle this issue, a number of recent works propose to unify reasoning and stream processing, giving rise to the research field of stream reasoning which supports decision systems based on the continuous processing of data streams together with rich background knowledge [8].

This paper presents a process which aims to use stream reasoning techniques in the medical domain in order to detect situations of interest based on data coming in real time from smart sensors. These situations of importance are defined using expert knowledge and may be used by a decision making system to raise an alarm depending on the health status of the patient monitored. To do so, an ontology representing the medical context of a person is defined featuring concepts such as constraints and situations [2] which enable to add a higher level meaning to the data. To detect these constraints and situations on the flow of data, logical queries are performed regularly according to the stream reasoning techniques. This paper also discusses the construction of these logical queries and asserts that the method presented is not limited to a certain type of diseases.

The article is structured as follows: Section 2 reviews existing works which propose methods to deal with data coming from IoT, especially in a medical context. Section 3 presents the ontology defined in this work to represent a patient under medical surveillance with IoT. Section 4 details the method developed to process the flow of data in order to detect situations of importance in real time and Section 5 shows an example of implementation of this process on real life temperature data and discusses the general application of this work. Finally, Section 6 serves as a conclusion.

2. Related Works

Large amounts of data provided by IoT need to be analysed to become useful. There already exists a lot of method from various domain to perform such kind of analysis, in particular in the context of health data [9]. In this domain, the challenge comes from the various type of smart sensors available, which may provide different types of data at different frequencies.

This section presents existing methods to deal with large amount of heterogeneous data and discusses the limitations of these approaches to provide a high-level analysis of medial data streams in real-time taking into account the patient's context.

2.1. Analysing health data

It exists multiple approaches to work on data in the context of health monitoring where decisions have to be made on noisy information. The underlying problem to solve is to identify the context in which the data are produced in order to build a meaningful interpretation.

Statistical techniques are commonly used to perform this task, ranging from hidden Markovian methods [10] to Bayesian networks [11] or conditional random fields [12]. All these methods rely on expert knowledge to create a network of states representing the patient health and context. Training data are used to compute the transition parameters of these networks. This training implies the need of a large amount of labeled data which represents an important drawback as it makes these techniques impossible to use on patient monitored on few days.

Machine learning techniques, from neural networks [13] to decision tree [14] or clustering methods as k-nearest neighbor [15], have been used in a medical context to estimate the state of a patient depending on data measured by sensors. This family of methods uses artificial intelligence techniques to create a classification between normal and abnormal behavior in large amount of data. By nature, machine learning techniques are able to combine heterogeneous sources of information and does not require an important quantity of expert knowledge as they learn to detect patterns by themselves. However, these techniques require a lot of training data before being able to give a relevant result. These methods also lack of explainability regarding the result they propose, making it difficult to interpret for a decision making system which should be trusted by the patient and the physician.

To solve the problems imposed by the need of training data, it exists statistical methods which only use a local focus on the data to identify trends [16] or to identify change points on time series [17]. These methods compute differences on a set of data to build a classification, either pairing values with a class of trend previously established or detecting change points on time series without assuming their meaning. These analysis methods show interesting results however they only work on numerical time series and do not take into account the context of measures (i.e. the analysis is carried out on raw data and does not adjust its results depending on the context of the measure). Moreover, these techniques only work offline as they require computation time and a broad understanding of the time series they are evaluating.

Finally, semantic approaches have been proposed to study the health of a patient monitored with smart sensors by abstracting data to higher level concepts easier to interpret. These methods includes fuzzy systems [18], logical rules [19] or ontologies [20] [21] and they all rely on expert knowledge from physicians with no need for training data. However, these approaches only work on offline static data where it is possible to take time to compute valid logical rules, which is not suitable to detect valuable situations on real time data.

2.2. Reasoning on a data stream

The challenge of performing logical reasoning on real time data lead to the development of Stream Reasoning [8]. It has already been used in a variety of area, including the monitoring of smart cities to improve the efficiency of services [22] [23], the semantic analysis of social media to improve the traditional analysis based on graph between users and activities [24] [25], the maritime safety and security to execute logical reasoning over ship trajectories [26], the adaptation of spatio-temporal behavior of robots depending on theirs own static knowledge [27], or in the context of telemedicine concerning the monitoring of specific diseases taking into account the context of the patient [28].

Stream reasoning techniques continuously perform logical queries in specific periods of time on data streams together with previous knowledge represented in an ontology. The idea is to confront the flow of data coming from sensors with a known representation of the state of the world, creating context for real time data.

It exists various stream reasoning engines [29] , C-SPARQL (Continuous SPARQL) [30] and CQELS [31] are two of the most known and used. C-SPARQL adds continuous query capabilities to the SPARQL query language over timestamped RDF streams with a periodic execution strategy. RDF streams are an ordered sequence of pairs where each pair is constituted by a RDF triplet and its timestamp `t : (<Subject,Predicate,Object>,t)`. C-SPARQL uses time windows to decompose data stream and deal with both RDF streams and static background knowledge, represented as RDF triples. On the other side, CQUELS also combines data streams with static knowledge but uses the arrival of new triples as a trigger for its queries.

Cascading reasoning is another method used to perform a reasoning on an ontology in real-time. It works by reasoning quickly on sub-parts of the ontology in real-time while executing complex request on the whole ontology in the background. This has been used on ambient-intelligent care rooms to triggers alarms depending on data produced by IoT [32]. The authors present a framework which enables to execute various layers of reasoning using fog computing [33] to improve the computation time of the system. However, this work does not deal with complex rules which may raise a preventive alarm as it only tries to trigger an alert as soon as possible when a problem on the patient happened.

2.3. Health Ontologies

Stream reasoning techniques work with a representation of the background knowledge of the studied domain; in a medical context, this implies to use an ontology to represent concepts linked to a patient, her medical environment and her disease while having a representation of sensors used for the monitoring.

The Saref4Health ontology [34] proposes to represent precisely sensors used in a medical context, being more particular than the classical SSN ontology [35] used to represent all kind of sensors. However, this ontology does not covers information about the patient, the disease or the patient's context while being very specific about medical sensors at the same time.

The notion of context for the person monitored has been already discussed on the case of quality of life for old people living in their home [36], [37]. These works use a semantic context around a sensor to interpret its value and detect the activity the person is performing [38]. However, in these works, the medical information of the patient is never considered and the context of a measure is limited to the values produced by other sensors in a smart house; it does not take into account events out of the sensor system.

It exists multiple ontologies related to the medical domain [39]. However, most of these work offer a very precise representations of molecular or biological information related to medicine or very specific diseases. None of them propose either a general representation of the concept of diagnostic nor the environmental context of a patient such as the every day activities which may impact the values produced by smart sensors.

Finally, there is a need for better interoperability between electronic health record representation (EHR) [40]. EHR are a numerical representation of an inpatient and all the data gathered during the stay at he hospital with currently multiple standards for their representation. The use of ontology could help to improve this interoperability but there are currently no such ontology.

2.4. synthesis

This section shows that none of the existing methods currently used for analysing health data makes it possible to work in real time, with heterogeneous sensors and taking into account the context of the patient. Without the patient's context, during the monitoring, it is then complicated to identify meaningful data in order to help a decision making system. The only statistical solutions usable in real time would require important quantity of training data which are complicated to obtain on patients monitored for only few days.

Semantic methods make it possible to overcome the lack of training data by using expert knowledge. This knowledge enables to identify weak signals which, once combined, mark the existence of a meaningful situation in the live monitoring of a patient. The definition of constraints and situations on an ontology has been previously carried out in the context of industry 4.0 for the preventive detection of machine failure [41]. The detection of constraints and situations over data streams while using static knowledge stored in an ontology enables to identify moments of interest with a semantic meaning. Then, these semantic meanings may be used to help a decision making system to raise an alert or not concerning the health of the monitored person.

This paper presents an ontology to represent a patient monitored and her context as well as a method to use this ontology in order to interpret the data streams coming from IoT. Stream reasoning techniques are used on time-windows to compute logical queries combining the flow of data with the static context of a patient. The ontology and the reasoning process developed use expert knowledge to help a decision making system work on sparse data, while the use of patient's context enables the system to work even with a defecting sensor.

3. Describing the patient's context with an ontology

In order to interpret values sent by sensors according to the context of a patient, it is mandatory to have a semantic representation of the concepts forming this context. This semantic representation enables to enrich the interpretation of raw data, taking into account, for example, the history of a patient or the activity performed at the moment of a measure. The purpose of the SICoPaD[1] ontology, presented in figure 1, is to represent a patient, the medical information related to this patient, the sensors used to monitor this patient, the measures produced by these sensors, the activities which may have an impact on these measures as well as abstract concepts used to analyse and sum up at a high level these data. This section defines the SICoPaD ontology while explaining how it has been developed by reusing well known ontologies and linking them with new components.

[1]The name of the ontology comes from the name of the project, SICoPaD, supporting this work.

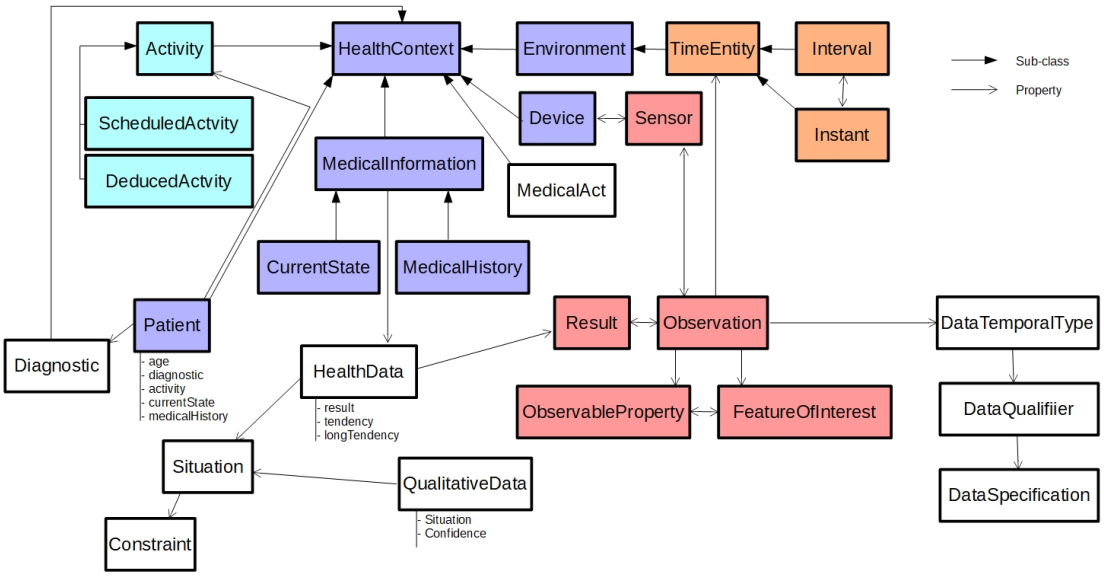

Fig. 1. Main concepts represented in the SICoPaD ontology.

3.1. Building the SICoPaD ontology with existing ontologies

The SICoPaD ontology integrates concepts from already existing ontologies to model the medical context of a patient monitored at home with IoT:

- The **SSN ontology** [35] (in red on figure 1) enables to represent a system of sensors and the values produced through the observation of specific properties. It is used here to gather information produced by smart sensors monitoring the medical state of a patient (temperature, blood pressure, heart rate, weight, etc.) as well as information concerning each sensor (name, type, brand, confidence level, etc.)
- The **Time ontology** [42] (in orange on figure 1) is used to define concepts related to duration and instants with a semantic meaning. This means the ontology may represent the time where a measure has been produced by a particular sensor, or the duration of an activity performed by a monitored patient, for example.
- *Guerma et al.* [43] (in blue on figure 1) propose an ontology to represent the medical context of a patient. This includes representing the personal information of a patient, the patient's medical information, present and past, but also the patient's environment and the devices used for the monitoring. All these notions are sub-classes of an abstract node, centralizing all the concepts related to the medical context and making the ontology modular.
- **CONON** (CONtext ONtology) [44] (in turquoise on figure 1) defines concepts related to general human context. Only the representation of human activities is kept here as the rest of the context surrounding the patient is defined by the work of *Guerma et al* [43]. The definition of activities enables to take into account what the person is doing in a day (eating, running, sleeping, etc.) which may influence values measured by sensors.

These four ontologies enable to define the necessary concepts to represent the context of a patient monitored at home which are : the patient and the patient's medical information, the activities she may perform, the sensors observing particular properties of the patient and a representation of time events. With these concepts, it is possible to integrate and perform logical queries on medical data depending on the time these data are gathered. However, it lacks some key concepts to interpret values send by sensor. Also, it is not enough to create complex logical queries in order to detect situations of interest for the potential raising of alarms.

3.2. Adding missing concepts to the SiCoPAD ontology

To complete the definition of medical context for a patient monitored at home and create link between concepts of interest, nine ontological classes (in white on figure 1) are added to the works presented in section 3.1 to create the SiCoPAD ontology:

- **Diagnostic**: description of the initial disease for which the patient is monitored. It may be used as an abstract concept to connect a more detailed ontology to describes diseases. Without any other ontologies connected to it, the initial diagnostic is described by an identifier according to the International Classification of Disease, Tenth Revision (ICD-10) [45].
- **Medical act**: representation of medical acts performed by or at the demand of a physician (taking a particular medicine, performing a certain action, etc.). It could be an abstract concept used to connect a more detailed ontology representing medical acts.
- **Health data**: generalization of values produced by sensors with a computation of short term and long term tendencies for each.
- **Qualitative data**: abstract concept used to associate a situation detected with a confidence level. This concept is made to be used by the decision making system.
- **Constraint**: abstract concept representing defined restrictions on certain properties of any concept in the ontology. Details are given in section 4.
- **Situation**: abstract state representing particular conditions of interest during the monitoring of a patient at home, depending on a combination of constraints. Details are given in section 4.
- **Data Temporal Type**: Definition of the temporal type of the values observed. These values may be punctual or continuous, indicating their frequency.
- **Data Qualifier**: Description of the type of values which may be produced; it may be qualitative or quantitative.
- **Data Specification**: Precision for the type of data generated by sensors, if it is a numerical value with or without unit, a percentage or a qualitative event.

The description of the initial diagnostic is added to the framework proposed by *Guerma et al* as it may modify the values or the interpretation of values produced by sensors. For example, different cancers at different phases may impact the evolution of body temperature [46]. With the same idea, a description of medical acts is added as it seems important to know if a patient already took some medicine or did something with the advice of a physician ; e.g. if someone took a medication to reduce body temperature, it is interesting to see if it produces the expected effect or not.

The definition of health data and qualitative data enables to centralise important information from the point of view of the decision making system. The main idea is to track back the source of an alarm raised by an external decision making system still keeping high-level semantic tokens, with the aim to keep the system explainable.

The Data Temporal Type, Data Qualifier and Data Specification concepts are used to describe precisely the type of data a sensor may produce regarding a particular observation. It is used to define, in advance, the logical rules used to detect situations, as explained in section 4. For example, a glucose meter produces punctual numerical values with a unit which implies a long time-window to gather multiple data and the possibility to check the belonging of the given value regarding a given range. A thermometer which produces a value every 30 seconds will be considered as generating continuous numerical data with a unit and will be used with a shorter time-window with possibility to detect the crossing of a given threshold. A smart pillbox which indicates if it is open or close is seen as creating punctual qualitative values with only two possible results which implies a bigger time-window and the comparison of the real value with an expected value.

Finally, constraints and situations represent a generalization of their initial definition as they were first introduce in the context of industry 4.0 [47]. They both enable to define static concepts aggregating knowledge from the ontology in order to detect meaningful moments in the desired context. These two concepts are detailed in section 4.

3.3. Linking resources together

All the modules presented in this section are integrated in the same ontological model with the help of relationships between concepts, added to the existing relationships coming from previous works described in section 3.1. Here are the main relationships which make the backbone of the SICoPaD ontology :

- **performsActivity** links the patient with the activity module, indicating what activity the patient is performing.
- **hasDiagnostic** links the patient to an initial diagnostic.
- **hasCurrentState** links the patient with a current medical state, containing multiple health data.
- **hasMedicalHistory** links the patient with a medical history which is composed of all the previous current states.
- **hasSensor** links the device concept with a sensor. The device concept may be aligned to the system concept from the SSN Ontology [35].
- **hasHealData** links the medical information abstract concept with multiple health data. It makes it possible for a current state or a medical history to feature some health data.
- **basedOnResult** links a health data with a result produced by a sensor about an observation.
- **hasTemporalType** links an observation with a data temporal type in order to describe in advance the type of value produced by a sensor over a given observation.
- **hasTime** links an observation with a time entity to describe the moment when an observation happened.

The main idea of the SICoPaD ontology, which can be seen with its relationships, is the modularity, based on the work of *Guerma et al.* [43]. This means sub-classes from the health context concept may act as abstract classes on which it is possible to connect more detailed ontologies to describe these concepts, such as the diagnostic, the patient or the medical acts. This is what has been done by aligning the device concept with the system concept from the SSN ontology or by connecting the Time Ontology to the environment concept. This modularity is eased by the definition of relationships between these main abstract classes.

3.4. Ontology evaluation

Ontology evaluation refers to the processing of assessing the expressiveness, accuracy, and quality of an ontology from the knowledge representation perspective. The aim of ontology evaluation is to avoid logical inconsistencies or undesired inferences [48]. To make sure the SICoPaD Ontology is free of modelling errors and anomalies we use the web-based ontology evaluation tool OOPS! (OntOlogy Pitfall Scanner!) [48]. As an automatic ontology pitfall detector, OOPS! helps to detect the most common pitfalls that appear during the ontology development process.

In OOPS, an ontology is evaluated on three dimensions: structural dimension, that focuses on mistakes detection on syntax and formal semantics; functional dimension, which considers the intended use and functionality of the proposed ontology; and usability-profiling dimension, evaluates the level of ease of communication when different users use the same ontology. In all the dimensions mentioned above, the possible ontology pitfalls are classified into three importance levels: critical, important, and minor. Critical pitfalls may affect the ontology consistency, reasoning and applicability. It is crucial to correct these pitfalls. Important pitfalls are not critical for ontology function but recommended to be corrected. Minor pitfalls are those that do not represent a real ontology engineering problem. However, removing minor pitfalls may help to better organise the ontology and improve its usability.

To evaluate the quality of the SICoPaD ontology, we uploaded the ontology to the web-based OOPS! platform [2]. As results, some minor pitfalls were detected. These ones do not affect the consistency, reasoning or applicability of the ontology. The main issue mainly regards ”missing domain or range” errors inherited from the SSN ontology. The documentation of SSN states that not adding the domain or range to certain properties is a design decision not to be restrictive with them. However, we have decided to add the missing range and domains. The final evaluation of the SICoPaD ontology by OOPS! appears in Figure 2.

[2] http://oops.linkeddata.es/

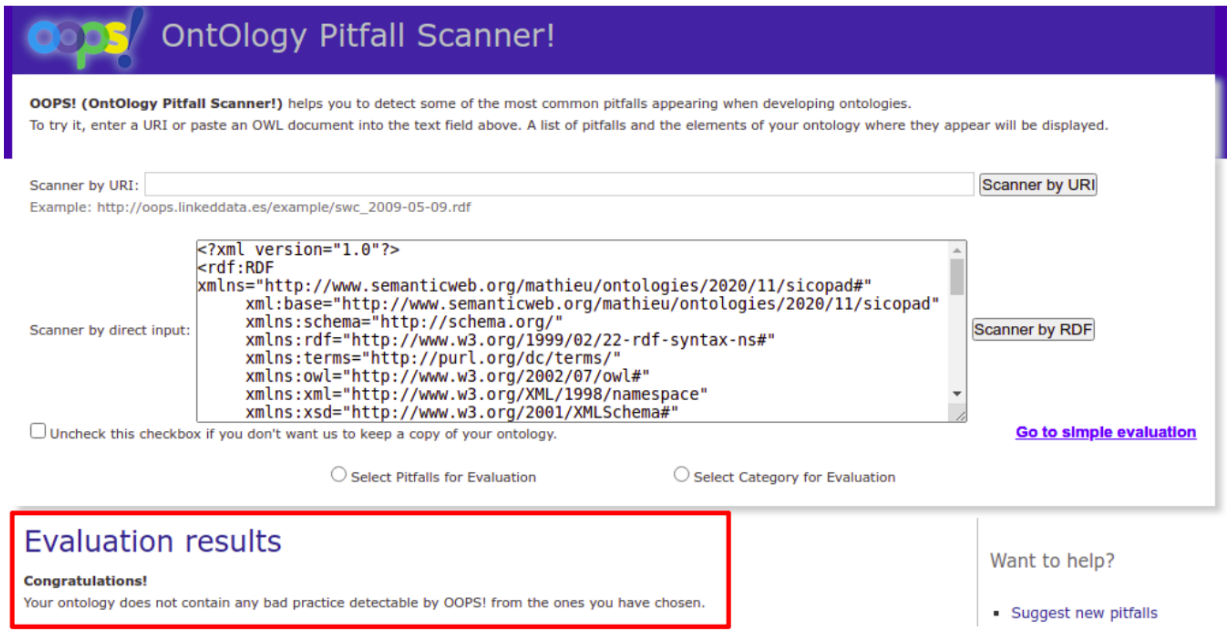

Fig. 2. Screenshot of the proposed ontology evaluation results by OOPS!

4. Using live data to detect situations of importance

The ontology presented in section 3 enables to describe the context of a patient monitored with smart sensors. This static context, compared to the flow of data, is used to detect situations of importance in real time. The goal of this process is to interpret the values measured by sensors in regard to the patient's static context. This section details the definition and usage of constraints and situations [47] based on expert knowledge and how to use these notions to detects moments of importance in real time.

4.1. Details about constraints and situations

A major issue when dealing with semantic technologies is the translation of expert knowledge into semantic concepts and rules. Constraints and situations concepts are introduced to ease this process when getting knowledge of importance from physicians for the description of complex medical states.

As defined in section 3, a constraint is an abstract concept representing defined restrictions on an observable property. Multiple constraints enable to define restrictions of interest without having any presupposed meaning to them; it may be a given value from a sensor (e.g. $bloodpressure = 240$), an interval of values for a sensor (e.g. $glycemia \in [0.8; 1.2]$), a threshold for a given value (e.g. $temperature > 38,5$), a given value, an interval or a threshold concerning the tendency of a sensor (e.g. $weightTendency \in [0.1; 0.2]$), a given activity (e.g. $activity = "sleeping"$), a particular medical act (e.g. $medicalAct = medicineTaken$), a disease the patient may suffer ($disease = diabetes$), a particular state for the patient (e.g. $age > 60$) or a temporal information (e.g. *happens the night* or *last 30 minutes*).

A situation is an abstract state representing a particular scenario of interest for the monitoring of a patient. Precisely, a situation is expressed as a combination of constraints and instances of situations from the ontology at the time of detection. In details, a situation may be described by a set of constraints (e.g. $hypoglycemia = glycemia < 0.8$; $highTemperatureNight = temperature \in [37.6; 37.8]$ & $time \in [00:00; 06:00]$ & *lasts at least 5 minutes*), a combination of constraints and previously detected situations (e.g. $fever = highTemperature$ & *lasts at least 30 minutes* ; $outOfFever = fever$ & $temperature \in [37.0; 37.4]$) or only a set of previously detected situations (e.g. $highFever = fever$ & $highBloodPressure$).

The definition of constraints and situations enables to ease the gather of expert knowledge as it transforms a statement from a physician into a set of abstract semantic concepts understandable by a computer. For example, if an expert says "the patient experiences fever if her temperature is above 38.0 degrees Celsius for at least 5 consecutive minutes", it translates into constraints $c_1 : temperature > 38,0$ and $c_2 : lasts\ at\ least\ 5\ minutes$ and into situation $s_1 : c_1 \ \&\ c_2$. This means that specialists can explain how they proceed when monitoring a patient without having to presuppose any high-level meaning to every type of observation they make. Their explanations can be very specific and personalized, decomposed into atomic parts (constraints) and reassembled into queries (situations). And as these instanced situations are part of the ontology, they can be reused to define more complex situations.

4.2. Description of the overall process

With stream reasoning techniques, queries are executed continuously over the data streams while considering an ontology which represents the current state of the world. An important point is to differentiate facts depending on the time they are added to the ontology as it makes a difference in the logical reasoning performed. At this extend, the concept of *current state* is introduced to consider differently a fact that is happening "right now" and a fact that happened previously.

The *current state* of a patient is the set of values produced by sensors monitoring the patient between two instants of logical reasoning; every measures produced by sensors and gathered by the system between two situation detection are forming the *current state* of the patient (i.e. what is considered to be the patient's present). The detection of situations is performed making the distinction between information which happened during the current state and facts which happened previously, which are considered as the medical history of the patient (or the static knowledge of the system). Moreover, values included in the *current state* keep their time stamp so it is possible to detect temporal relations among them.

The overall process proposed in this article is described by algorithm 1 with t_0 the starting time of the current state, t the current time, and δt the duration of the time window representing the current state. The process works as follows: while the patient is monitored, values collected by sensors are gathered to create the current state of the patient until the next set of logical queries is executed. At the end of the time-window, the system identifies constraints which are satisfied during the current state and, based on these constraints and using the ontology, detects situations. Finally, data that is no longer needed is removed from the ontology (an operation called "clean medical history" in algorithm 1) while the newest data acquired during the current state is put into the historical context, and the process repeats by starting a new *current state* with a new time-window.

```
while true do
    while t < t_0 + δt do
        gather data from sensor;
    end
    Identify constraints;
    Detect situations;
    Clean medical history;
    Add the gathered data to medical history;
    t_0 ← t;
end
```

Algorithm 1: Overall process to detect situations of importance with stream reasoning

The parametrization of the overall process is carried out with expert knowledge and knowledge about the type of sensor, as described in section 3, used for each situation detection. The δt parameter in algorithm 1 represents the length of the time-window associated with a particular situation detection. It may be different depending on the type of sensor, the type of data gathered or situation detected. For example, with a thermometer that gives a value every 30 seconds, it makes sense to have time-windows of few minutes to detect fever while with a blood

pressure sensor that collects 4 values a day, it seems better to have larger time-windows to detect abnormal behavior. This customisation is also possible for the definition and refinement of constraints and situations depending on the patient; the thresholds or intervals used to define constraints are not fixed for every patient monitored and may be adjusted depending on expert knowledge. The global aim is to adapt to the context of the monitoring, i.e. the patient and the various sensors in use.

5. Example case and discussion

The ontology defined in section 3 has been implemented in OWL [49] and the process presented in section 4 has been implemented with the C-SPARQL engine [30]. These two parts of the current work are tested using two example cases which are described in this section. The goal of the first example is to show the coverage of the SICoPaD ontology while the goal of the second example is to show the detection of situations running on a real-life sample of temperatures from patients under monitoring.

5.1. Coverage of the ontology

Showing the coverage of the SICoPaD ontology means showing it is able to represent all the concepts required to model a patient monitored by smart sensors. Accessing real Electronic Medical Records (EMR), which compile day-to-day data for inpatients, is complicated and/or costly. To overcome these difficulties, EMR generated with EMRBots [50] are used for this example case.

EMRBots create random EMR files from which it is possible to pinpoint the type of information related to a patient that is worth noting. These meaningful information are the identity of the patient, the initial diagnostic and a series of time-stamped medical values for each inpatient generated. The representation of the identity of a patient and the representation of the initial diagnostic are obviously covered by the SICoPaD ontology according to the requirements expressed in section 3. The representation of medical time-stamped values is carried out with the use of the SSN [35] and the Time [42] ontology which are part of the SICoPaD ontology. In order to identify constraints and detect situations over these values, it is necessary to know if they are part of the current state of the patient or part of the medical history of the patient, as explained in section 4. To do so, each medical value is represented by the *Result* concept from the SSN ontology. Each *Result* is linked to a new instance of the *HealthData* concept thanks to the *basedOnResult* relation. This *HealthData* instance is linked to an abstract state represented by the *MedicalInformation* concept thanks to the *hasHealthData* relation. This *MedicalInformation* state is first stored as a *CurrentState* and then as a *MedicalHistory* as explained in section 4.

Figure 3 shows an illustration of this coverage for a patient randomly generated by EMRBots [50] identified by identification number 1A8791E3-A61C-455A-8DEE-763EB90C9B2C. The second time this person was hospitalized, his initial diagnostic was "Underdosing of coronary vasodilators" which correspond to the reference T46.3X6 in the International Classification of Disease, Tenth Revision (ICD-10) [45]. Among the various variables measured during the second stay at the hospital, the glucose was measured at least two times, which is represented on figure 3 by the fact that the oldest measure of 73.6 mg/dL is part of the medical history of the patient, while the newest measure of 129.8 mg/dL is part of the current state, each time through the use of the abstract *HealthData* concept.

5.2. Detection of situations on a real-life case

The detection of situations of interest must be realized on real life data, where real situations happened. The Henri Becquerel hospital center in Rouen, France, carried out a medical test on the use of thermometric caps to inspect their efficiencies in front of more traditional thermometers. The thermometric caps send a temperature value from the inside of a patient every 30 seconds, while traditional methods only produce measures every 4 hours. This case study consists of data concerning three patients monitored during 3 to 7 days. Among these data, there are two cases of fever (temperature > 38.0 degrees Celsius) which are situations of interest for the system.

The first example is a patient who is experiencing a temperature above 37.6 degrees Celsius during the night (usually the body temperature is lower in the morning and higher in the evening, about 37.2 degrees Celsius during

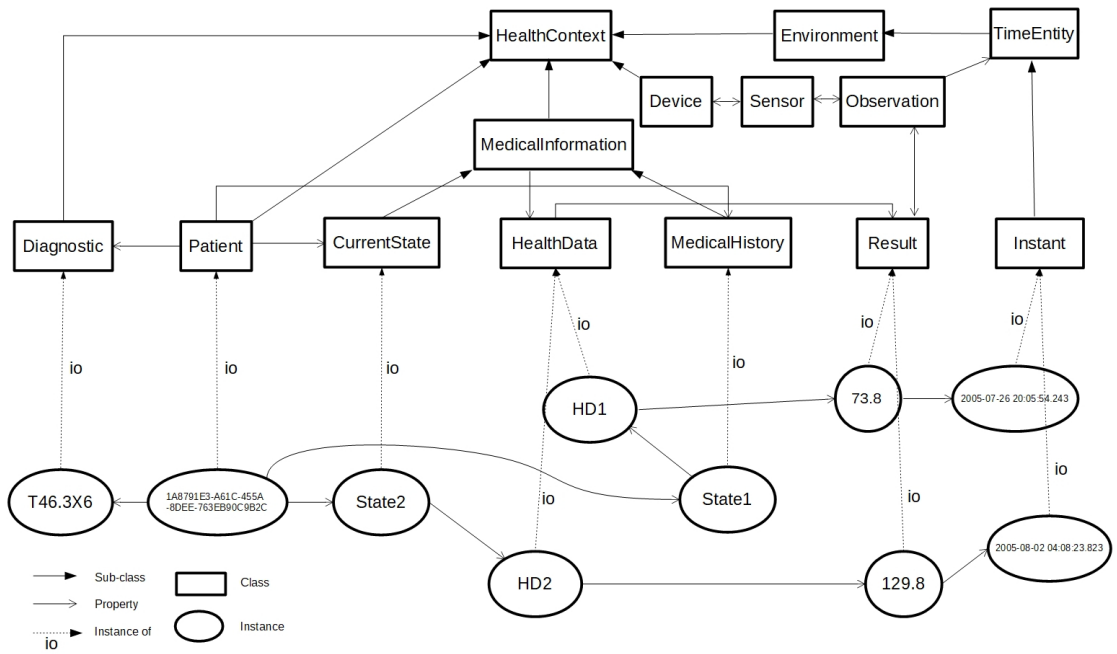

Fig. 3. Example of a patient monitored by smart sensors represented with the SICoPaD ontology

the day) which leads to very high temperatures the next morning (above 37.8 degrees Celsius) before going above the 38.0 degrees Celsius threshold. Finally, the temperature goes back to normal, supposedly because the patient took medicine, according to experts. The second example is a patient who is experiencing a very high fever (temperature above 38.5 degrees Celsius) during a whole day. This example enables to detect pre-fever state (temperature $\in [37.8; 38.0]$) and also to show the difference between fever and high fever.

Discussions with physicians to understand the cases enabled to establish a list of constraints and situations, respectively in table 1 and 2. Constraints on the temperature values are in degree Celsius and make it possible to create different intervals. Constraints on time enable to indicate the moment some other constraints are satisfied (especially during the night) and avoid taking into account worthless values. The five situations defined for these use cases represent moments of interest related to the temperature of the patient, according to physicians.

Table 1
Definition of constraints

Constraint ID	Description
c1	temperature $\in [37.0; 37.6]$
c2	temperature $\in [37.6; 37.8]$
c3	temperature $\in [37.8; 38.0]$
c4	temperature $\in [38.0; 38.5]$
c5	temperature > 38.5
c6	happens between 00:00 and 06:00
c7	last at least 5 minutes
c8	last at least 30 minutes

Table 2
Definition of situations

Situation ID	Definition	Description
s1	no other situation detected	normal situation
s2	c2 & c6 & c7	high temperature during the night
s3	c3 & c7	pre-fever
s4	(c4 & c7) or (s3 & c8)	fever
s5	c5 & c7	high fever

As explained in section 4, and more precisely with algorithm 1, the overall process works as follows on the example data: during 20 minutes (the time-window), the system collects temperature values sent by the sensor in real time, one every 30 seconds, to build up the "current state" of the patient; queries are executed at the end of the 20 minutes to detect situations based on the identified constraints given by physicians; once situations are raised for

this time-window, the current state is transferred to the patient's medical history and the process starts all over with the gathering of new data during a new time-window.

The C-SPARQL query presented in Listing 1 has the purpose of detecting the "high temperature during the night" situation (s2). The query name is registered on line 1 and prefixes used in the query are declared on lines 2 and 3. The query is executed on RDF streams that correspond to the property temperature (`S_temp`) in the time frame of 20 minutes, sliding the window by 20 minutes (line 5). The chosen time frame is arbitrary and can be changed as desired. It produces triplets of values (line 4): the sensor name (`?s`), the patient (`?p`) and the average temperature during that period of time. In order to obtain the patient to which the sensor is attached, we indicate in the query that the C-SPARQL engine must use our ontological model as background knowledge (line 6). Line 8 enables to obtain the patient to which the sensor is attached. To get the observation's values, `?o1` individual is bound with the data values `?v1` through the appropriate properties (`sosa:madeObservation` and `sosa:hasSimpleResult`) (line 9-12). Finally, the list of output triplets are filtered out to include only the ones where the observations satisfy the restrictions in the `FILTER` and the `HAVING` clauses (line 13-18). In other words, the observations whose timestamps of the observations are between 00h and 06h (c6), and the average temperature is between 37.6 and 37.8 (c2).

```
1   REGISTER QUERY PrefievreNuit AS
2   PREFIX : <http://semanticweb.org/Sicopad#>
3   PREFIX sosa: <http://www.w3.org/ns/sosa/>
4   SELECT ?s ?p (AVG(?v1) as ?avgTemp)
5   FROM STREAM <Stream_S_temp> [RANGE 20m TUMBLING]
6   FROM <http://streamreasoning.org/sicopad>
7   WHERE {
8      ?p  sosa:hosts           ?s .
9      ?s  sosa:madeObservation ?o1 .
10     ?o1 sosa:hasSimpleResult ?v1 .
11     ?o1 :hasTime             ?t  .
12     ?t  :inXSDDateTime       ?ts .
13  FILTER (
14     (hours(?ts) >= 0) &&
15     (hours(?ts) < 6) )
16  GROUP BY ?s ?p
17  HAVING ((AVG(?v1) >= 37.6) &&
18          (AVG(?v1) <= 37.8)) .
```

Listing 1: C-SPARQL query to detect the high temperature during the night situation.

5.3. Results and discussions

The coverage example detailed in this section shows that only few concepts defined in the SICoPaD ontology were necessary to represent the electronic medical record of a generated patient. For more clarity, Figure 3 only shows a sample from all the information possible to represent on this use case, but it is not a problem to represent information on the sensors or on the various typology of data produced to monitor the patient.

As the patient has been randomly generated for this example, it implies that the SICoPaD ontology is able to completely represent the meaningful information, such as the identity, the initial diagnosis and all the medical values, with a time stamp, to track the evolution of the overall medical state concerning any inpatient. Furthermore, with the remaining concepts in the SICoPaD ontology, it is possible to model a medical act performed on the patient or describe an activity the patient would have done. These information are not present on an EMR but it could be of importance for the high level analysis of medical values like the detection of situations.

The use of constraints and situations concepts is shown with the second example case where it is used to detect particular situations on different sets of temperature values with only two sets showing relevant events (i.e. the patient developed a fever). A graphical representation of the situations detected while streaming the data in real-time may be found in figure 4 and figure 5. The first example describe a situation of high temperature during the night which leads to fever in the morning while the second example shows the situation of a fever turning into a high fever. The vertical lines represent the instants where a situation has been detected, with a label indicating which

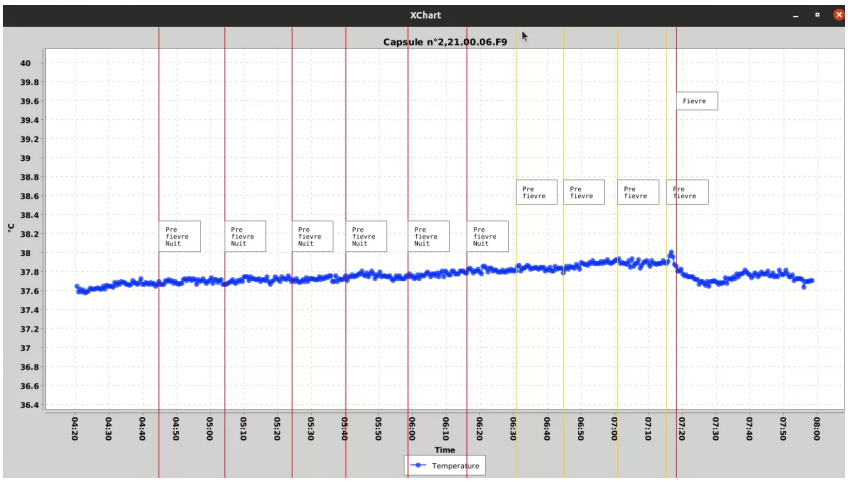

Fig. 4. Graphic representation of the situations detected in real-time on a first set of data

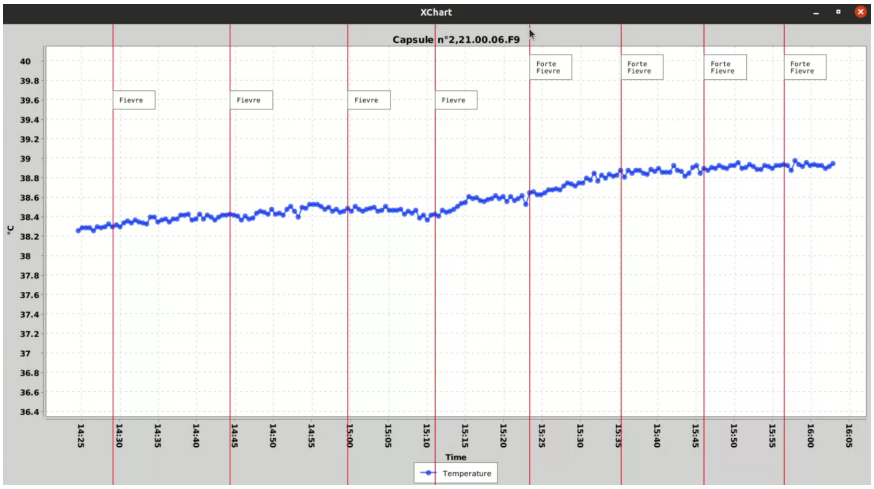

Fig. 5. Graphic representation of the situations detected in real-time on a second set of data

situation has been raised. The time-window for this detection was 20 minutes. Videos of the examples and source code may be found at: https://gitlab.insa-rouen.fr/fgiustozzi/sicopad

Results observed on figure 4 and figure 5 show the system correctly identified situations, defined on table 2, when they happen: the first example raises situation *s2* during the end of the night, then situation *s3* when the temperature goes above 37.8 degrees Celsius before triggering the situation *s4* because the patient stayed in situation *s3* for more than 30 minutes; the second example first raises situation *s3* (not shown on figure 5) before detecting situation *s4* because the temperature is above 38.0 degrees Celsius and finally detecting situation *s5* because the temperatures is above 38.5 degrees Celsius. It is also notable that, when tested on other part of these data, no situation other than the normal situation (*s1*) were detected, and that all these detections are realized in real time on the flow of data coming at a rate of a value every 30 seconds.

This second use case represents a proof of concept of the process described in section 4. The detection of these type of situations may easily be used by a decision making system to alert, in real time, physicians of a problem the patient is experiencing. The main advantage with using context is to add a meaning to a flow of raw data which may be complicated to analyse in real-time, even by an expert. However, the goal stays the combination of various kind of data coming in real-time from different sensors and the complexification of constraints and situations so a

set of weak signals out of the norm may help the raise of a correct preventive alarm, thus helping to take in charge a patient having troubles before what is done currently. However, this complexification could not have been tested due to the small amount of detailed data available.

Finally, the values fixed in this example for the intervals of temperature, the duration to observe a phenomenon or the length of the time-window, were chosen after discussions with experts on the particular cases studied. However, it is possible to adapt them to different patients or other flows of values; this adaptation must be realized with the advice of experts so the situations detected are really of interest.

6. Conclusion

This article presents an ontology to describe the context of a patient monitored with smart sensors and a process to use this ontology in real-time, giving a semantic meaning to a flow of data. This flow of data is represented as a RDF stream that is integrated with the ontology detailed in section 3. Stream reasoning methods are then used to detect meaningful situations based on previously identified constraints, as explained in section 4. These situations are easier to understand than the raw flow of data and may then be used by a decision making system. The coverage of the ontology is shown in section 5 while the overall process is tested and discussed on a proof of concept use case using two patients who make fever episodes, which are detected in real-time on real life data.

This work is part of a project which aims to alert physicians when the health of a patient monitored at home is deteriorating. This implies that the situations detected by the method described in this paper are meant to be sent to a decision making system which decides the type of alert to raise depending on a combination of situations. The results presented in section 5 show the proposed method is able to detect situations prior to a dangerous moments, thus helping a decision making system to raise an alert in advance of critical times. The use of semantic technologies such as ontologies and stream reasoning ensures the explainability of the decision made by the system, thus strengthening the confidence a physician may have on the overall system.

For the future, the process must be tested on more complex examples with multiple sensors, more constraints and situations defined and more information for each patient (medical case, blood pressure, weight, and glycemia). The goal of these tests on more complex cases is to detect more complex situations using at the same time various medical data flow and concepts represented in the ontology. The final result should be able to identify the real situation in a given context with a suitable confidence level in order to help a decision making system to raise not too much nor not too few alerts, and then help physicians for the monitoring of patients at home. The system may also put in evidence situations of interest currently not known by physicians, and thus help providing more accurate treatment to patients.

Acknowledgement

This work is part of the SICoPaD Project financed by Region Normandie from France

References

[1] L. Atzori, A. Iera and G. Morabito, The internet of things: A survey, *Computer networks* **54**(15) (2010), 2787–2805.
[2] F. Giustozzi, J. Saunier and C. Zanni-Merk, Context modeling for industry 4.0: An ontology-based proposal, *Procedia Comp. Science* (2018).
[3] A. Zanella, N. Bui, A. Castellani, L. Vangelista and M. Zorzi, Internet of things for smart cities, *IEEE Internet of Things journal* (2014).
[4] S.R. Islam, D. Kwak, M.H. Kabir, M. Hossain and K.-S. Kwak, The internet of things for health care: a comprehensive survey, *IEEE access* **3** (2015), 678–708.
[5] W. Jin and D.H. Kim, Design and implementation of e-health system based on semantic sensor network using IETF YANG, *Sensors* **18**(2) (2018), 629.
[6] T.R. Gruber, A translation approach to portable ontology specifications, *Knowledge acquisition* **5**(2) (1993), 199–220.
[7] E. Della Valle, S. Ceri, F. Harmelen and D. Fensel, It's a Streaming World! Reasoning upon Rapidly Changing Information, *Intelligent Systems, IEEE* **24** (2010), 83–89.

[8] H. Stuckenschmidt, S. Ceri, E. Della Valle and F. Van Harmelen, Towards expressive stream reasoning, in: *Dagstuhl Seminar Proceedings*, Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2010.

[9] H. Mshali, T. Lemlouma, M. Moloney and D. Magoni, A survey on health monitoring systems for health smart homes, *International Journal of Industrial Ergonomics* **66** (2018), 26–56.

[10] A.R.M. Forkan, I. Khalil, Z. Tari, S. Foufou and A. Bouras, A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living, *Pattern Recognition* **48**(3) (2015), 628–641.

[11] D. Ge, W. Li and L. Wei, An intelligent reasoning machine for the home-based care, in: *2015 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, IEEE, 2015, p. S–9.

[12] D.J. Cook, Learning setting-generalized activity models for smart spaces, *IEEE intelligent systems* **2010**(99) (2010), 1.

[13] X. Xie, J. Guo, H. Zhang, T. Jiang, R. Bie and Y. Sun, Neural-network based structural health monitoring with wireless sensor networks, in: *2013 Ninth International Conference on Natural Computation (ICNC)*, IEEE, 2013, pp. 163–167.

[14] L. Gao, A. Bourke and J. Nelson, Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems, *Medical engineering & physics* **36**(6) (2014), 779–785.

[15] V. Jakkula, Predictive data mining to learn health vitals of a resident in a smart home, in: *Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007)*, IEEE, 2007, pp. 163–168.

[16] W.W. Melek, Z. Lu, A. Kapps and W.D. Fraser, Comparison of trend detection algorithms in the analysis of physiological time-series data, *IEEE transactions on biomedical engineering* **52**(4) (2005), 639–651.

[17] C. Truong, L. Oudre and N. Vayatis, Selective review of offline change point detection methods, *Signal Processing* **167** (2020), 107299.

[18] E. Shimokawara, T. Kaneko, T. Yamaguchi, M. Mizukawa and N. Matsuhira, Estimation of basic activities of daily living using zigbee 3d accelerometer sensor network, in: *2013 International Conference on Biometrics and Kansei Engineering*, IEEE, 2013, pp. 251–256.

[19] W. Jih, J.Y.-j. Hsu, C.-L. Wu, C.-F. Liao and S. Cheng, A multi-agent service framework for context-aware elder care, in: *AAMAS'2006 Workshop on Service-Oriented Computing and Agent-Based Engineering (SOCABE'2006)*, 2006, pp. 61–75.

[20] K. Wongpatikaseree, M. Ikeda, M. Buranarach, T. Supnithi, A.O. Lim and Y. Tan, Activity recognition using context-aware infrastructure ontology in smart home domain, in: *2012 Seventh International Conference on Knowledge, Information and Creativity Support Systems*, 2012.

[21] R. Shojanoori and R. Juric, Semantic remote patient monitoring system, *Telemedicine and e-Health* **19**(2) (2013), 129–136.

[22] S. Tallevi-Diotallevi, S. Kotoulas, L. Foschini, F. Lécué and A. Corradi, Real-time urban monitoring in dublin using semantic and stream technologies, in: *International Semantic Web Conference*, Springer, 2013, pp. 178–194.

[23] F. Lécué, S. Kotoulas and P. Mac Aonghusa, Capturing the pulse of cities: Opportunity and research challenges for robust stream data reasoning, in: *Semantic Cities@ AAAI*, 2012.

[24] P. Mika, Flink: Semantic web technology for the extraction and analysis of social networks, *Journal of web semantics* **3**(2–3) (2005), 211–223.

[25] G. Erétéo, M. Buffa, F. Gandon and O. Corby, Analysis of a real online social network using semantic web frameworks, in: *International semantic web conference*, Springer, 2009, pp. 180–195.

[26] G.M. Santipantakis, A. Vlachou, C. Doulkeridis, A. Artikis, I. Kontopoulos and G.A. Vouros, A stream reasoning system for maritime monitoring, in: *25th International Symposium on Temporal Representation and Reasoning (TIME 2018)*, 2018.

[27] D. de Leng and F. Heintz, Ontology-Based Introspection in Support of Stream Reasoning, in: *Thirteenth Scandinavian Conference on Artificial Intelligence (SCAI), Halmstad, Sweden, 5-6 November 2015*, IOS Press, 2015, pp. 78–87.

[28] J.-P. Calbimonte, J.-E. Ranvier, F. Dubosson and K. Aberer, Semantic representation and processing of hypoglycemic events derived from wearable sensor data, *Journal of ambient intelligence and smart environments* **9**(1) (2017), 97–109.

[29] D. Dell'Aglio, E. Della Valle, F. van Harmelen and A. Bernstein, Stream reasoning: A survey and outlook, *Data Science* **1**(1–2) (2017), 59–83.

[30] D.F. Barbieri, D. Braga, S. Ceri, E.D. VALLE and M. Grossniklaus, C-SPARQL: a continuous query language for RDF data streams, *International Journal of Semantic Computing* **4**(01) (2010), 3–25.

[31] D. Le-Phuoc, M. Dao-Tran, J.X. Parreira and M. Hauswirth, A native and adaptive approach for unified processing of linked streams and linked data, in: *International Semantic Web Conference*, Springer, 2011, pp. 370–388.

[32] M. De Brouwer, F. Ongenae, P. Bonte and F. De Turck, Towards a cascading reasoning framework to support responsive ambient-intelligent healthcare interventions, *Sensors* **18**(10) (2018), 3514.

[33] A.V. Dastjerdi, H. Gupta, R.N. Calheiros, S.K. Ghosh and R. Buyya, Fog computing: Principles, architectures, and applications, in: *Internet of things*, Elsevier, 2016, pp. 61–75.

[34] J. Moreira, L.F. Pires, M. van SINDEREN and L. Daniele, SAREF4health: IoT Standard-Based Ontology-Driven Healthcare Systems., in: *FOIS*, 2018, pp. 239–252.

[35] A. Haller, K. Janowicz, S.J. Cox, M. Lefrançois, K. Taylor, D. Le Phuoc, J. Lieberman, R. García-Castro, R. Atkinson and C. Stadler, The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation (2019).

[36] L. Wang, T. Gu, X. Tao, H. Chen and J. Lu, Recognizing multi-user activities using wearable sensors in a smart home, *Pervasive and Mobile Computing* **7**(3) (2011), 287–298.

[37] Q. Zhang, M. Karunanithi, R. Rana and J. Liu, Determination of activities of daily living of independent living older people using environmentally placed sensors, in: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2013, pp. 7044–7047.

[38] L. Chen, C.D. Nugent and H. Wang, A knowledge-driven approach to activity recognition in smart homes, *IEEE Transactions on Knowledge and Data Engineering* **24**(6) (2011), 961–974.

[39] N.M. Meenachi and M.S. Baba, A survey on usage of ontology in different domains, *International Journal of Applied Information Systems* **4**(2) (2012), 46–55.
[40] E. Adel, S. El-Sappagh, S. Barakat and M. Elmogy, Ontology-based electronic health record semantic interoperability: A survey, in: *U-Healthcare Monitoring Systems*, Elsevier, 2019, pp. 315–352.
[41] F. Giustozzi, J. Saunier and C. Zanni-Merk, Abnormal situations interpretation in industry 4.0 using stream reasoning, *Procedia Computer Science* **159** (2019), 620–629.
[42] J.R. Hobbs and F. Pan, Time ontology in OWL, *W3C working draft* **27** (2006), 133.
[43] H. Guermah, T. Fissaa, H. Hafiddi, M. Nassar and A. Kriouile, Context modeling and reasoning for building context aware services, in: *2013 ACS international conference on computer systems and applications (AICCSA)*, IEEE, 2013, pp. 1–7.
[44] X.H. Wang, D.Q. Zhang, T. Gu and H.K. Pung, Ontology based context modeling and reasoning using OWL, in: *IEEE annual conference on pervasive computing and communications workshops, 2004. Proceedings of the second*, Ieee, 2004, pp. 18–22.
[45] N.C. for Health Statistics et al., International classification of diseases, tenth revision (ICD-10), 2009.
[46] M.-C. Mormont and F. Levi, Cancer chronotherapy: principles, applications, and perspectives, *Cancer: Interdisciplinary International Journal of the American Cancer Society* **97**(1) (2003), 155–169.
[47] F. Giustozzi, STEaMINg: semantic time evolving models for industry 4.0, PhD thesis, Normandie, 2020.
[48] M. Poveda-Villalón, A. Gómez-Pérez and M.C. Suárez-Figueroa, OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation, *International Journal on Semantic Web and Information Systems (IJSWIS)* **10**(2) (2014), 7–34.
[49] D.L. McGuinness, F. Van Harmelen et al., OWL web ontology language overview, *W3C recommendation* **10**(10) (2004), 2004.
[50] U. Kartoun, Advancing informatics with electronic medical records bots (EMRBots), *Software Impacts* **2** (2019), 100006.
[51] C.S. de Oliveira, F. Giustozzi, C. Zanni-Merk, C. Sanin and E. Szczerbicki, Stream Reasoning to Improve Decision-Making in Cognitive Systems, *Cybernetics and Systems* **51**(2) (2020), 214–231.
[52] O.W. Group et al., OWL 2 Web Ontology Language Document Overview: W3C Recommendation 27 October 2009 (2009).