Toward Situation Awareness for the Semantic Sensor Web: Complex Event Processing with Dynamic Linked Data Enrichment

Souleiman Hasan, Edward Curry, Mauricio Banduk, Sean O'Riain

Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland
{souleiman.hasan, ed.curry, mauricio.banduk, sean.oriain}@deri.org

Abstract. Over the past few years there has been a proliferation in the use of sensors within different applications. The increase in the quantity of sensor data makes it difficult for end users to understand situations within the environments where the sensors are deployed. Thus, there is a need for situation assessment mechanisms upon the sensor networks to assist users to interpret sensor data when making decisions. However, one of the challenges to realize such a mechanism is the need to integrate real-time sensor readings with contextual data sources from legacy systems. This paper tackles the data enrichment problem for sensor data. It builds upon Linked Data principles as a valid basis for a unified enrichment infrastructure and proposes a dynamic enrichment approach that sees enrichment as a process driven by situations of interest. The approach is demonstrated through examples and a proof-of-concept prototype based on an energy management use case.

Keywords: situation awareness, semantic sensor networks, semantic web, linked data, dynamic enrichment, complex event processing, spreading activation, semantic similarity.

1 Introduction

The notion of Situation Awareness (SA) has emerged in two main fields: Information Fusion [1] and Human Computer Interaction [2]. The objective of situation awareness is to empower the user with an understanding of the developing relationships of interest between entities in question within a specific time and space [1]. SA techniques have been applied to improve user understanding within a range of systems, from the mission and safety critical role of helping pilots in the cockpit, to empowering business executives' with decision support to optimize business operations with real-time business intelligence [1].

As sensor networks deployments have increased, sensor information has become one of the main information flows within situation awareness systems. At the same time the introduction of web and semantic web technologies to sensor networks [3] has improved the accessibility of sensor data [4]. Enterprises are also finding more uses for sensors, from supporting the operational layers to the higher-level strategic decision making layers [5].

Sensor readings are usually limited in the amount of information they hold. The quality of SA is dependent on the quality of context available when the situation awareness is determined. Thus, there is a need to enrich sensor information flows with additional context from existing systems within the enterprise to conduct higher quality situational assessments.

In this paper we investigate the challenges associated with situation awareness in web sensor networks, we propose an approach to situational awareness utilizing a combination of Complex Event Processing (CEP) and Linked Data. In particular, we examine the validity of using linked data as a basis for sensor data enrichment. The approach utilizes dynamic enrichment over linked data streams, combined with CEP as the means to realize situation awareness in enterprises.

The contribution of this paper is the introduction of dynamic enrichment as a key enabler to realize situation awareness over large-scale and open web sensor networks. The paper proposes a model for dynamic enrichment based on spreading activation in linked data and the semantic similarity measures between information items and the situations of interest. It also proposes an evaluation framework for the approach.

The remainder of the paper goes as follows: *Section 2* motivates the need for a situation awareness mechanism for the web sensor networks along with some associated challenges. *Section 3* describes the proposed approach and details the dynamic enrichment process. *Section 4* demonstrates the approach via a prototype based on an energy management use case. *Section 5* summarizes briefly related work in situation awareness and enrichment. The paper concludes in *Section 6* with future directions.

2 The Need for Situation Awareness for Web Sensor Networks

Over the last few years there has been a proliferation in the use of sensors within different use cases, from air and water pollution monitoring, to machinery health monitoring within factories. The increased uptake is being driven by lower costs to buy and install sensors and the simplification of their deployment [4]. The indications are this trend is set to continue with the introduction of web-based open standards for sensor networks and the switch to open data licensing policies which will further increase the accessibility of sensor data [4].

Within business environments there is an increasing demand to support real-time decision making business process. When making a decision the value of information is higher and more useful for the decision makers when its freshness is higher [6]. This motivates the desire to expand the use of sensor networks upward within the knowledge and decision stacks of enterprises, from supporting technical low-level applications, to supporting higher-level decision making processes [5]. Nevertheless, users and organizations find it hard to interpret, understand and leverage the rapidly increasing quantity of information which necessitates the use of situation assessment mechanisms.

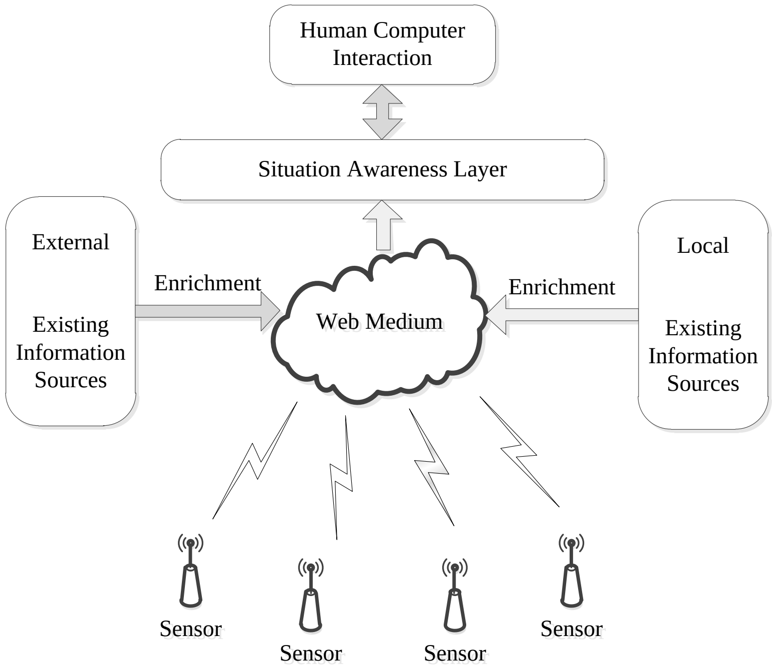

Fig. 1. A situation awareness layer positioned upon the sensor web

Situation awareness has been defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" [7]. Elements of the environment include people, projects, devices, rooms, etc. Sensor networks can provide an enterprise with (near) real-time fresh flows of information items (i.e. readings, observations, events, etc.). The synergy between these dynamic information flows along with traditional data sources that contain rather static information increases the quality of the overall comprehension of relationships between elements in the enterprise (e.g. people, devices, rooms, products, etc.) and thus the quality of business status assessment.

In order to process their information flows, many enterprises employ systems that are dedicated to high rate information flow processing in addition to their traditional database management systems. Data Stream Management Systems (DSMS) and Complex Event Processing (CEP) systems have been adopted with commercial systems starting to appear in the last few years [8]. *Figure 1* illustrates how a situation awareness layer can be positioned upon web sensor networks to deliver higher-level insights.

2.1 Challenges with Situation Awareness

The process of creating situation awareness requires the configuration of the underlying information systems to process raw information flows and abstract them up to the level of situation awareness; refer to *example 1*. Within current state-of-the-art of DSMS and CEP systems, SA configuration is done by defining patterns of information flow items that are mapped to situations of interest for the target users [8]; refer to *code snippet 1* which shows an implementation of the scenario exposed in *example 1*.

Example 1.
Within an energy management scenario sensors observe the *kWh* energy usage of 12 heaters distributed among 3 floors in a building. Motion detection sensors are also in place to detect if a floor is empty. Typically, observing the energy consumption of devices and the emptiness of a floor does not provide in itself much value. That is because of the granularity level that might be non-useful for users who are not in the operational level and because these observations need to be linked together and drawn against other contextual information to make the result more actionable from an energy saving perspective. It would be better if after detecting that a floor was empty, the energy usage observations were aggregated over the devices in that floor for a time period (e.g. 30 minutes) and then compared with an acceptable threshold in order to conclude a more useful piece of information such as an excessive energy usage. That allows the users to move from a massive amount of data to higher level knowledge and facilitates the decision making with regard to energy saving.

To express the scenario explained in *example 1* in a pattern language such as the Event Processing Language (EPL) used in the open source complex event processing engine Esper [9], the following expression is used (simplified):

```
INSERT INTO ExcessiveEnergyUsageByFloor
SELECT a.floor as floor
FROM PATTERN [(a=FloorEmptySensor -> every
         b=DeviceEnergyUsageSensor(a.floor=b.floor))]
    .WIN:TIME(30 min)
GROUP BY a.floor
HAVING SUM(b.usage) > GetAcceptableThreshold(a.floor)
```

Code Snippet. 1. EPL implementation of the scenario explained in *example 1*

The following challenges can be identified along with the different activities needed for situation assessment:

Bridging the Information Gap.
One of the main challenges with defining SA is the need to bridge information gaps between different levels in an enterprise (e.g. operational to strategic) [5]. In technical terms this means defining patterns of interest in languages close to SQL (moving from `FloorEmptySensor` and `DeviceEnergyUsageSensor` to

`ExcessiveEnergyUsageByFloor` in *code snippet 1*), or sometimes, using user interfaces to help construct the patterns from known information flows and a controlled vocabulary [10]. This can become extremely challenging within open and large-scale environments, and even more difficult at web-scale. That is due to the large number of possible patterns and the large number of information flows and items' properties to be considered in patterns.

Heterogeneity of Information Flows.

Another difficulty results from the heterogeneous usage of ontologies, i.e. terminology or vocabulary, to publish semantic sensor data by different publishers. This complicates the task of the person responsible for defining the situations; the situation manager. It becomes very difficult to integrate terms from different publishers on a web-scale.

Uncertainty about Occurrence and Content of Information Flow Items.

Some real-world events might not be observed or vice-versa. Errors might also occur in the content of sensor readings. This results in a degree of uncertainty about what really happens in the real-world and affects the definition and evaluation of situations of interest. For instance, exact matching between situations of interest and observations could result in unfavorable false positives and false negatives.

Putting Information Flows into Context.

Sensor readings are usually limited in the amount of data that they contain (refer to *example 2*). This can be due to the limited resources of sensors and also the scope of the environment the sensor can observe. When used within an enterprise, sensor readings will often need to be interpreted within the context of other information systems including Enterprise Resource Planning (ERP), financial accounting systems, energy management systems, etc. Thus, the amount of data the item contains should be expanded in order to include information relevant to more situations of interest. This is a process known as data enrichment. Enriching sensor data adds further complexity as it can be difficult to define in advance and must be maintained during the system lifetime.

A more extensive discussion on challenges in situation assessment can be found in [11]. In the following we focus more on the enrichment issue.

3 Situation Awareness for Semantic Sensor Networks

In order to realize situation awareness for information flows from sensor networks and existing systems within the enterprise, we propose the use of a Complex Event Processing engine along with a dynamic enrichment component that enriches the information items before they can be considered for evaluation; refer to *figure 2*.

A loose control over the systems and information flows is assumed due to large-scale and openness motivated by the adoption of web technology. Thus, there is a need for a unified enrichment mechanism. The use of sensor networks that respect linked data principles [12] when publishing data forms a solid basis for enrichment. URIs can be used to refer to related entities in the enterprise and open linked data cloud. The sensor data can be enriched with useful information such as RDF data that is retrieved when dereferencing a URI; see *example 2*.

Beyond the concept of enrichment, we propose the idea of dynamic enrichment where the enrichment strategy is decided at run-time and depends mainly on the semantic similarity between the situations of interest registered in the system and the attributes of observed information items. Dynamic enrichment brings the following benefits:

- It simplifies the integration of context data into SA systems and thus simplifies the definition of situations of interest;
- Dynamism allows SA systems to quickly evolve;
- Semantic similarity reduces the gaps between different vocabularies used to describe items;
- Web data (external) sources can be easily included (weather data, partner information such as power mix of an electricity supplier, etc.)

Figure 2 illustrates the suggested approach to reach situation awareness in semantic sensor networks with more focus on the dynamic enrichment component. It is further explained in the following sub-sections.

3.1 Complex Event Processing

A Complex Event Processing [5] engine provides the processing model for evaluating situations of interest. After the situation of interest is expressed in the configuration of the CEP engine in the form of an event pattern, new information items can participate in the evaluation of the pattern if they are relevant. When a pattern is matched, a new higher-level event (e.g. `ExcessiveEnergyUsageByFloor` in *code snippet 1*) is generated and can participate in further processing or could be forwarded to an event consumer like a dashboard or a business process management tool.

3.2 Dynamic Enrichment with Linked Data

In order to address the challenge of defining and maintaining enrichment strategies for distributed and heterogeneous information flows, there is a need to support dynamic event enrichment. That means that enrichment is not defined during the design time of the system but left to the run-time where each information flow item is enriched according to different criteria; especially the situations of interest that are defined. *Figure 2* illustrates the main steps and factors that affect the proposed dynamic enrichment process. We will consider *examples 1* and *2* as well as *code snippet 1* while we are walking through the proposed approach.

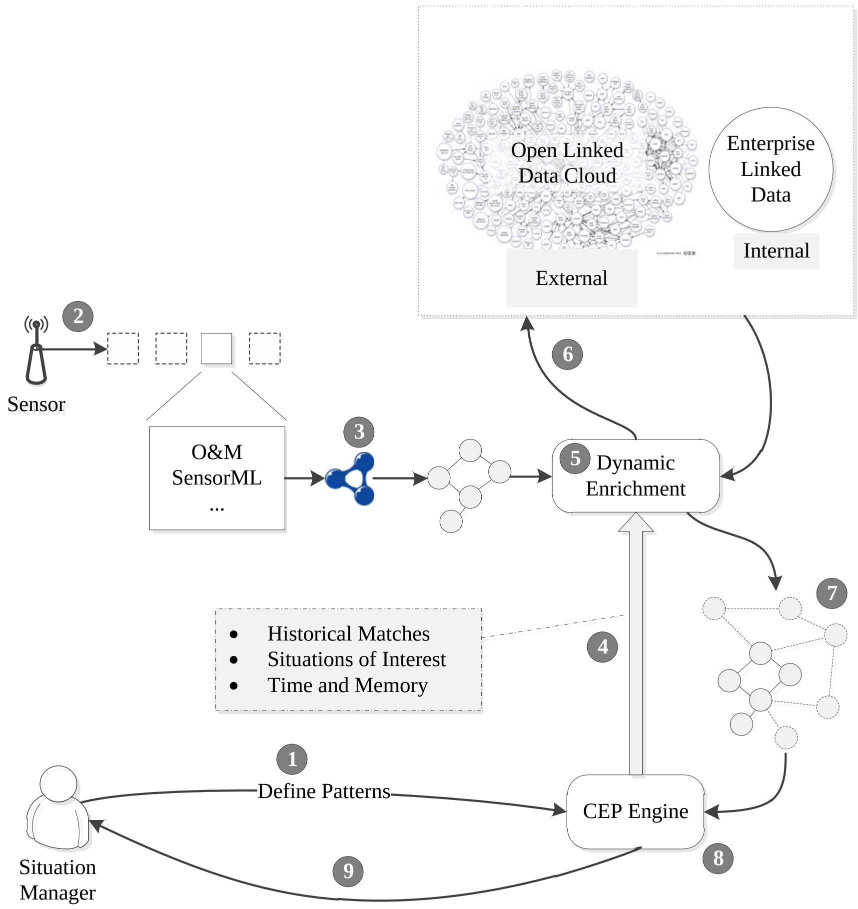

Fig. 2. Dynamic enrichment of Linked Sensor Data. The situation manager defines the situations of interest in the *CEP engine* (1), the sensor data are produced in *O&M* and *SensorML* formats during the run-time (2) then converted to RDF according to linked data principles (3), the *Dynamic Enrichment* component takes into consideration factors from the *CEP engine* such as historical matches, situations of interest, time and memory available (4) and decides on the data and time for enrichment (5), the enrichment is done by a spreading activation over the linked data graphs (6) and results in enriched sensor data (7) which is then evaluated against situations of interest (8), matches are forwarded to the end user (9).

Within this approach information items are adapted to linked data near the sensors with URIs referring to existing data entities in the enterprise or on the web of data. For example, the sensor readings of a heater' energy usage might come out of the sensors in an O&M XML format [13] containing the IP address of the sensor with the amount of energy usage. The linked data adapter converts these messages to an RDF format like N3 [14] and replaces the IP address by the appropriate URI of the heater in question. The resulting message would look like the one in *example 2*. More best practices about publishing linked sensor data can be found in the literature [15].

Example 2.

In *code snippet 1*, the `DeviceEnergyUsageSensor` reading may include just one RDF triple that describes the sensor observation about a specific device. The triple would use the URI of the device which would return more information about that device such as its type or the floor it is installed in when it is dereferenced. The sensor reading triple would look like the following:

```
<http://energy.deri.ie/resource/device/H008070>
<http://energy.deri.ie/ontology#usage> 55.6.
```

After the linked sensor readings reach the enrichment component, the component determines the information items, amount, and time for data enrichment. Enrichment itself is done by spreading activation [16] over the linked data graph starting from the content of information items. The direction and amount of spreading activation is guided by the semantic similarity between information items and the situations of interest. Spreading activation over linked data has been used for different purposes; see [17] as an example of spreading activation use for natural language querying over linked data. *Figure 3* shows an example of spreading activation for the `DeviceEnergyUsageSensor` reading.

We propose the following criteria as a basis for the enrichment decision:

- A semantic similarity measure between the information item content and the potential situation patterns that the information item can participate in. For example, the situation of interest in *example 1* is concerned with the accumulated energy usage of heaters that are installed in a *floor*. The sensor reading does not have data about the floor where the heater is installed (*example 2*). Semantic similarity is then used to guide the spreading activation process until satisfactory information about the heater's floor is found. That might take one dereferencibility step for the URI `<http://energy.deri.ie/resource/device/H008070>` to find a predicate `<http://rooms.deri.ie/ontology#installed>` that leads to a resource of type `<http://rooms.deri.ie/ontology#Floor>`;
- The amount of time and memory available for the CEP engine to meet the user need to deliver the situation awareness in time. In order to improve performance effective caching is important. For example when we get *N* readings about the energy consumption of the same device, the device URI should be dereferenced in the first time and the result kept in the cache for the following times. The lifetime of an item in the cache should depend on a probabilistic or stochastic model that predicts the occurrence of events in the future;

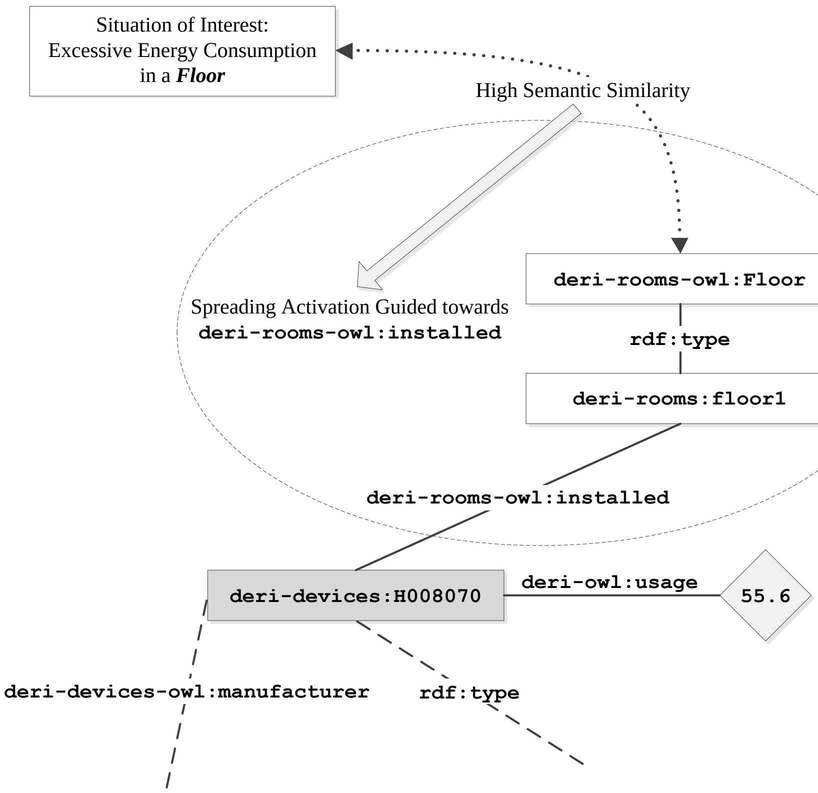

Fig. 3. Spreading activation for the `DeviceEnergyUsageSensor` reading

- The knowledge about useful previous enrichment or non-useful previous enrichment from the perspective of matched situations. For example, if the message in example 2 was enriched with the manufacturer of the device but it has never been used for matching, so there is no need to enrich with it the next time.

Table 1 summarizes the relationship between different criteria and the decision of enrichment.

Table 1. Criteria affecting different dimensions of the enrichment in the proposed approach

	Items for Enrichment	Direction of Enrichment	Amount of Enrichment	Time of Enrichment
Semantic Similarity	Yes	Yes	-	-
Response Time to conduct SA	Yes	-	Yes	Yes
Available Memory for CEP engine	Yes	-	Yes	Yes
Previous Matches	Yes	-	-	Yes

4 Proof of Concept: Energy Management Use Case

In order to support the argument made throughout this paper, a proof-of-concept prototype has been developed based on an enterprise energy management use case. The use case builds on the examples covered in the previous sections. This section briefly covers the technicalities and experience while implementing the scenarios.

In a typical modern office building there are many sources of power consumption such as Heating, Ventilation and Air Conditioning (HVAC) systems, lights and electronic devices. Tracking the operation of these systems can help in identifying information related to energy leaks and non-ecological actions. This information can be utilized to achieve reductions in energy consumption and cost saving. The purpose of a building energy management system is to gather data related to energy consumption and to present it in an actionable manner where actionable implies minimal effort to move from the presented knowledge to energy-related decisions.

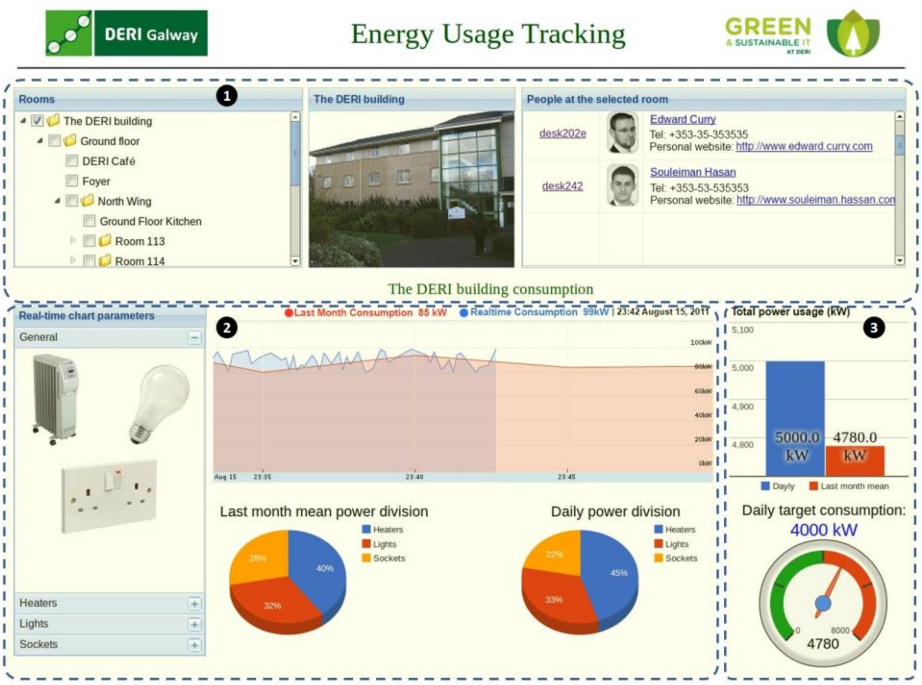

Fig. 4. A screenshot of the system dashboard. In (1) reference data from the enterprise linked data cloud can be seen; it is used for enrichment in the scenarios, in (2) instant measures by the sensors are shown and in (3) situation awareness is achieved by comparing the accumulative consumption with historical usage data and usage targets to detect high usage situations.

The system is deployed in the DERI office building. The information passing through the system is produced by 31 fixed energy consumption sensors covering office space, café, data centre, kitchens, conference and meeting rooms, computing museum along with 5 mobile sensors for devices, light and heaters' energy consumption as well as motion detection. Observations are collected by the sensor controller which triggers a broadcast of the information received. The sensor readings are adapted to RDF using the Jena framework [18] and enriched based on the enterprise linked data cloud that exists in DERI, which was developed in a previous project (see [19] for more information about Sustainable DERI project). The data is then sent to the CEP Engine. The CEP Engine makes situation assessment based on the pre-defined patterns of interest and once new data is generated by the engine it is forwarded to the user interface; refer to *figure 5* as an example screenshot.

To put the proposed approach into practice, basic energy usage sensor readings are sent without appropriate context information, such as in which floor or room of the building the consuming device is installed. A set of patterns of interest that aggregate energy usage according to the floors and rooms are registered in the CEP engine. The dynamic enrichment component does the necessary enrichment to include the missing pieces of information and allow the readings to be included in the evaluation of the deployed patterns. The system works as expected but a systematic evaluation is underway to evaluate the approach; see *Section 6*.

The CEP engine was extended to accept linked data events. Nevertheless, the core processing model is still a relational query model. This issue has not been investigated yet as we are more concerned with the enrichment part not with the matching functionality. However we believe that a deeper change in the processing model of the CEP engines is needed in order to effectively process Linked Sensor Data. We think that extending CEP with a more relaxed and approximate matching that is based on information retrieval approaches is more suitable for web deployments [20].

5 Related Work

Situation assessment has been identified as a key function in the Joint Directors of Laboratories (JDL) data fusion model [21]. It has been approached by different techniques ranging from probabilistic [22] to rule-based approaches [23]. Complex Event Processing (CEP) is a rule-based tool for processing dynamic information flows to help in situation assessment.

Sensor networks started to adopt semantic web technology in response to large-scale and heterogeneous deployments [3]. As a result, there has been a need to adapt CEP in order to process the semantic sensor web data [24]. Recently, situation awareness has been identified as one of the key challenges for semantic sensor networks [25]. Some works suggest the use of logic-based reasoners over RDF streams [24] but challenges such as performance and handling of uncertainty exist with such approaches in real-world scenarios [11];

Enrichment for information flows has been considered as a typical pattern in Message-Oriented Middleware (MOM) [26]. However, it has been considered as an external task used along with channel bandwidth considerations. We are not aware of research work that tackles the enrichment problem as a standalone problem in itself. However, the problem has been recognized in the event processing community as a main future research challenge [27].

6 Conclusions and Future Work

This paper discussed the synergy between information coming from semantic sensor networks together with existing information sources in enterprises to achieve high quality situational awareness to support decision making process. We argue the need for dynamic enrichment of information flows as a practical approach in large-scale and open systems. We also show how semantic sensor networks that respect linked data principles form a valid basis for dynamic and unified enrichment. We demonstrated a proof-of-concept prototype from the energy management world.

Future work would include the evaluation of the dynamic enrichment approach. Evaluation will be conducted towards: fewer amounts of memory usage and short time for enrichment as well as high precision and recall measures of matched situations. While the current work is concerned with a generic extension of CEP engines to do the enrichment, another future direction will examine the processing models of the CEP engines in order to realize natural language and approximate matching of situations over semantic sensor data.

7 Acknowledgements

This work has been funded by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2).

8 References

1. Waltz, E., Llinas, J.: Multisensor data fusion. Artech House (1990).
2. Endsley, M.R., Garland, D.J.: Situation awareness: analysis and measurement. CRC (2000).
3. Lefort, L., Henson, C., Taylor, K., Barnaghi, P., Compton, M., Corcho, O., Garcia-Castro, R., Graybeal, J., Herzog, A., Janowicz, K., Neuhaus, H., Nikolov, A., and Page, K.: Semantic Sensor Network XG Final Report, W3C Incubator Group Report (2011). Available at http://www.w3.org/2005/Incubator/ssn/XGR-ssn/.
4. Sheth, A., Henson, C., Sahoo, S.S.: Semantic sensor web. IEEE Internet Computing. 12, 78-83 (2008).
5. Luckham, D.: The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems. Addison-Wesley Professional (2002).
6. Hackathorn, R.: Current practices in active data warehousing. DMReview. (2002).
7. Endsley, M.R.: Design and evaluation for situation awareness enhancement. In: Human Factors and Ergonomics Society Annual Meeting. pp. 97-101 (1988).
8. Cugola, G., Margara, A.: Processing flows of information: From data stream to complex event processing. ACM Computing Surveys Journal. (2011).
9. Esper Complex Event Processing Engine, http://esper.codehaus.org/.
10. Etzion, O., Niblett, P.: Event Processing in Action. Manning Publications Co. (2010).
11. Blasch, E., Kadar, I., Salerno, J., Kokar, M.M., Das, S., Powell, G.M., Corkill, D.D., Ruspini, E.H.: Issues and challenges in situation assessment (level 2 fusion). Journal of Advances in Information Fusion. 1, 122-139 (2006).

12. Linked Data: Design Issues, http://www.w3.org/DesignIssues/LinkedData.html.
13. Botts, M., Percivall, G., Reed, C., Davidson, J.: OGC sensor web enablement: Overview and high level architecture. GeoSensor networks. 4540, 175-190 (2008).
14. Notation 3, http://www.w3.org/DesignIssues/Notation3.
15. Barnaghi, P., Presser, M., Moessner, K.: Publishing Linked Sensor Data. In: The 3rd International Workshop on Semantic Sensor Networks (2010).
16. Collins, A.M., Loftus, E.F.: A spreading-activation theory of semantic processing. Psychological review. 82, 407 (1975).
17. Freitas, A., Oliveira, J., O'Riain, S., Curry, E., da Silva, J.: Querying Linked Data Using Semantic Relatedness: A Vocabulary Independent Approach. In: Natural Language Processing and Information Systems. 40-51 (2011).
18. McBride, B.: Jena: A semantic web toolkit. Internet Computing, IEEE. 6, 55-59 (2002).
19. Curry, E., Hasan, S., Hassan, U. ul, Herstand, M., O'Riain, S.: An Entity-Centric Approach to Green Information Systems. In: The 19th European Conference on Information Systems (ECIS) (2011).
20. Curry, E., Chambers, D., Lyons, G.: Extending message-oriented middleware using interception. In: Third International Workshop on Distributed Event-Based Systems. p. 32 (2004).
21. Steinberg, A.N., Christopher L. Bowman: Revisions to the JDL Data Fusion Model. Handbook of Multisensor Data Fusion (2001).
22. Das, S., Grey, R., Gonsalves, P.: Situation assessment via Bayesian belief networks. Information Fusion, 2002. In: The Fifth International Conference on. pp. 664-671 (2002).
23. Matheus, C.J., Baclawski, K.P., Kokar, M.M.: Derivation of ontological relations using formal methods in a situation awareness scenario. In: Proceedings of SPIE. pp. 298-309. Citeseer (2003).
24. Taylor, K., Leidinger, L.: Ontology-Driven Complex Event Processing in Heterogeneous Sensor Networks. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., and Pan, J. (eds.) The Semanic Web: Research and Applications. pp. 285-299. Springer Berlin / Heidelberg (2011).
25. Taylor, K., Ayyagari, A., De Roure, D.: Research Topics in Semantic Sensor Networks. Proceedings of the Semantic Sensor Network Workshop, p. Preface (2006).
26. Hohpe, G., Woolf, B., Brown, K.: Enterprise integration patterns. Addison-Wesley (2004).
27. Hinze, A., Sachs, K., Buchmann, A.: Event-based applications and enabling technologies. In: The Third ACM International Conference on Distributed Event-Based Systems - DEBS'09. 1 (2009).