A Semantic Approach for Process Annotation and Similarity Analysis

Tobias Weller(✉)

Institute AIFB, Englerstr. 11, 76128 Karlsruhe, Germany
tobias.weller@kit.edu
http://www.aifb.kit.edu

Abstract. Research in the area of process modeling and analysis has a long-established tradition. There are quite few formalism for capturing processes, which are also accompanied by a number of optimization approaches. We introduce a novel approach, which employs semantics, for process annotation and analysis. In particular, we distinguish between target processes and current processes. Target process models describe how a process should ideally run and define a framework for current processes, which in contrast, capture how processes actually run in real-life use cases. In some cases, current processes do not match the target process models and can even overhaul them. Therefore, one is interested in the similarity between the defined target process model and current processes. The comparisons can consider different characteristics of processes such as service quality measures and dimensions. Current solutions perform process mining methods to discover hidden structures or try to infer knowledge about processes by using specific ontologies. To this end, we propose a novel method to capture and formalize processes, employing semantics and devising strategies and similarity measures that exploit the semantic representation to calculate similarities between target and current processes. As part of the similarity analysis, we consider different service qualities and dimensions in order to determine how they influence the target process models.

Keywords: Process annotation · Similarity analysis · Process analysis · Quality of Service · Semantic process modelling

1 Introduction

Process modeling and analysis has multiple application domains. I.e. clinical pathways are an evidence-based response to specific problems and care needs in clinics. They support physicians by providing recommendations on the sequence and timing of actions necessary to achieve an efficient treatment of patients [1,2]. Each clinic has their own pathways based on their individual evidence and experience. Therefore, there are multiple pathways that target different problems and care needs [3–5].

H. Sack et al. (Eds.): ESWC 2016, LNCS 9678, pp. 883–893, 2016.
DOI: 10.1007/978-3-319-34129-3_57

However, physicians are not strictly restricted to the published pathways. Therefore, the process, defined in the pathway (target process model), can differ from the actually performed workflow (current process). As a result, there might be discrepancies between the published clinical pathways and the actually performed workflow, which is based on the decisions of the physician on how to treat the patient.

This situation is aggravated by the fact that there is a lot of data, generated and used during the treatment of patients, that needs to be managed and interpreted. In order to ease this task, the information can be captured semantically and used for comparisons. For this purpose, there are already many ontologies in the medical domain, which can be used to structure the semantic information – the Disease Ontology[1] that provides descriptions and related medical terms about human diseases and the Foundational Model of Anatomy (FMA)[2], which describes classes, structures and relationships of all parts of the human body. In addition, processes can be compared based on different service qualities and dimensions, such as complexity, runtime, outcome or costs.

The problem of having current processes that diverge from the defined target process models does not occur only in the medical domain. The same difficulties arise also in enterprises and in the domain of Internet of Things (IoT) applications, in which the actual communication flow between devices can diverge from a defined target process model. This is precisely the topics that we want to explore. One aspect is that it is debatable whether the current process performs better, in terms of certain service qualities and dimensions, than the defined target process model. Knowing the deviation and different outcome of the service qualities and dimensions could lead to the incentive of adapting the target process model.

Given a set of current processes and its target process model, we are interested in calculating the similarity between them, in order to be able to quantify the variety and see how different processes behave in terms of different service quality aspects and dimensions. The revelation of the effect of service qualities and dimensions can be used for different aspects. One aspect is to provide a confidence interval for a service quality variable. Another aspect could be the hint of adapting the target process model, if the current process instances diverge too much from the target process model. The adaption of the target process model can than be performed in respect to certain service qualities and dimensions.

2 State of the Art

An important aspect, in order to have a common point of view on processes, is to define the term *process*. We use the process definition from ISO 9000:2015 [6], which is given in the following.

[1] http://disease-ontology.org.

[2] http://sig.biostr.washington.edu/projects/fm/index.html.

Definition 1. *ISO 9000:2015 Process: Set of interrelated or interacting activities that use inputs to deliver an intended result.*

Note 1 to entry: Whether the "intended result" of a process is called output, product or service depends on the context of the reference.

Note 2 to entry: Inputs to a process are generally the outputs of other processes and outputs of a process are generally the inputs to other processes.

This definition is specifically related to quality management systems, but we aim to use it in a broader way. We do not focus on quality management systems in particular but rather on processes in general.

Process semantic annotations: There are already widely used ontologies, such as Dublin Core Schema[3] that provide a set of metadata that can be used to annotate resources. We can use these ontologies to annotate process elements like e.g. tasks and gateways of the target process model. The advantage of such schemata is that they can be integrated easily in order to annotate resources and provide interoperability with further datasets.

Semantic process-based formalizations and conformance checking: Business Process Abstract Language (BPAL) provides a formal semantic to process modeling languages [7,8] and allows enriching it with semantic annotations. The formal definition allows a verification of the used properties and the ontology-based annotations. Thus, this approach can be used to semi-automatically map current process instances to its target process model and verify the mapping according to a correct semantic annotation [9–12]. There are also some ontology-based annotations for process models available that can be reused [13,14]. In addition to semantic annotations, there are also ontologies available to describe the components of a process and the relationships between them such as SUPER [15,16] and the Process Specification Language[4], which has been approved as an international standard [17].

Service qualities and dimensions: Existing approaches describe how service qualities and dimensions can be captured [18]. Thereby, frameworks like SERVQUAL can be used to measure the quality of processes [19]. Service qualities from e-services [20] or other process performance indicators [21] can also be used as metrics to measure the performance of processes.

Process matching: There are a number of different process similarity measurements for comparing processes. Some uses node similarity, structural similarity, behavioral similarity and language based matching [22,23]. However, most of them focus on business processes [24,25] and do not distinguish between target and current process models. The similarity of processes is, among others, used to cluster processes [26].

Adaption of target process models: Approaches like e.g. Process Mining try to reveal hidden structures and create a target process by using i.e. log files

[3] http://dublincore.org.

[4] http://www.mel.nist.gov/psl/.

or other data produced by process instances [27,28]. These approaches reveal hidden structures but not the influence of processes on different service qualities and dimensions. However, process mining techniques can be used to discover new insights based on a created reference model from the current process data [29].

3 Problem Statement and Contributions

We focus on performing similarity analysis between target and current processes by exploiting the semantics of processes. The semantics that we use to compare process models consist, among others, of the semantic annotations (like labels and descriptions) that we add to the process models, a domain hierarchy of the process elements, the user roles (for example, only specific users are allowed to perform a task or a decision) and rules that define the workflow of processes. Based on the presented motivation and the current state-of-the-art, we formulate the following research question and its subquestions:
How do we benefit from the combination of process models with semantics in order to improve processes by performing similarity analysis?

- **RQ1** How can we formally specify process data with semantics?
- **RQ2** Which service qualities and dimensions can we use to compare processes?
- **RQ3** Which methods can we use to perform similarity analysis of target processes and current process data?

During the PhD we will develop an approach to annotate process data with semantic information and perform similarity analysis of target process models and current processes. This approach will be modeled in a common way, so it is generally applicable. In the following, we discuss the subquestions in more detail.

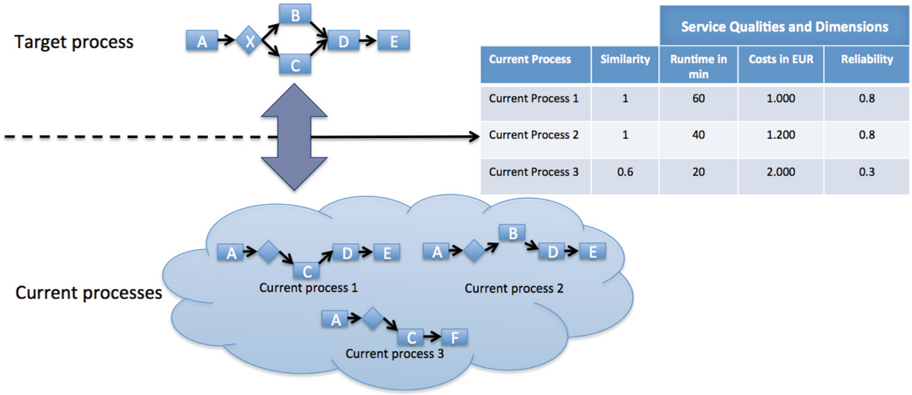
(RQ1) How can we formally specify process data with semantics?
There are already established formal representations for modeling languages e.g. for BPMN 2.0, the standard language BPMN 2.0 XML published by OMG[5] or the Petri Net Markup Language [30] for representing petri nets. However, while the execution semantics of processes is partly covered, there is a lack of semantics for the inputs/outputs used in the processes, annotations and the terminology of process elements. Therefore, we will show how to combine formally specified process models with semantics that can be queried and processed. The enriched current process instances can be used for comparisons and analysis.

(RQ2) Which service qualities and dimensions can we use to compare processes?
Processes can be compared based on different service qualities and dimensions such as runtime, outcome, costs or reliability. Capturing these service qualities and dimensions is a first step towards being able to compare the defined target

[5] http://www.omg.org/spec/BPMN/2.0/.

process model and the current processes. We will analyze existing frameworks (see Sect. 2) according to their extent and their usability in different domains and maybe extend them. As possible output, we may propose a new framework.

(RQ3) Which methods can we use to perform similarity analysis of target processes and current process data?
We will show which methods can be used to compare a target process model with a set of current processes. During the use of different similarity methods, we will exploit semantics such as the hierarchical arrangements of process elements, as well as domain semantics, and user roles, linked to tasks, and rules, which influence the process flow. Figure 1 shows the comparisons of a target process model with current process instances.



Current Process	Similarity	Service Qualities and Dimensions		
		Runtime in min	Costs in EUR	Reliability
Current Process 1	1	60	1.000	0.8
Current Process 2	1	40	1.200	0.8
Current Process 3	0.6	20	2.000	0.3

Fig. 1. Determining the similarity between target process model and current process instances

The research questions aim to result in multiple contributions. The first contribution is the introduction of an approach that integrates processes with semantic information that can be queried and processed. We would like to integrate as much semantic information as possible to allow, in a later step, enhanced similarity analysis that considers all these aspects. Another contribution is a set of service qualities and dimensions that can be used to compare processes. We will show different metrics and how they can be used in multiple domains. The last contribution is the similarity analysis between target process models and current processes. Thereby, we will use methods that exploit the semantics, captured in the previous step, such as the hierarchy of activities and process flows, to quantify the similarity.

4 Research Methodology and Approach

The structure of the research methodology and approach is directly derived based on the research questions (Sect. 3). Research methodologies can be classified as quantitative, qualitative and mixed research methodologies. Quantitative research methods collect numerical data and use it to analyze and explain a circumstance [31]. We will apply quantitative methodologies to plan and approach the research problems. In particular we will collect a sample of process instances, formalize the data, map them to the target process model, calculate the similarity between them and reveal the effect of current processes to service qualities and dimensions. We will investigate how semantics affects the analysis and comparisons of processes, and test different methods to compare processes.

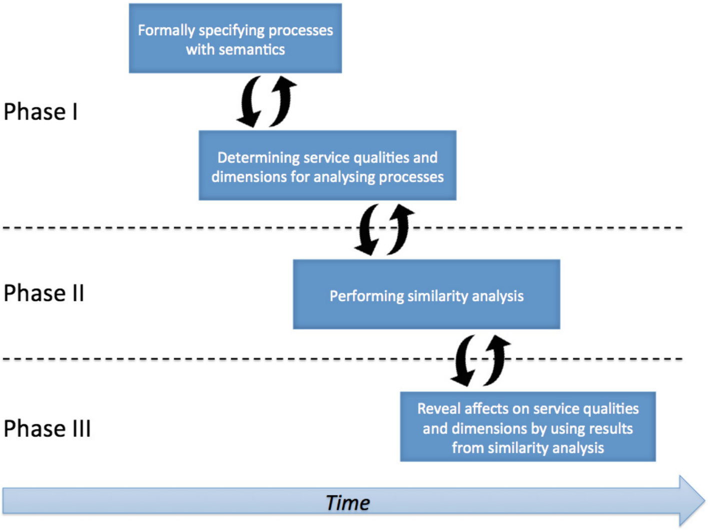

Phase I
Formally specifying processes with semantics
Determining service qualities and dimensions for analysing processes
Phase II
Performing similarity analysis
Phase III
Reveal affects on service qualities and dimensions by using results from similarity analysis
Time

Fig. 2. Research approach – divided into three phases. Each tackles another aspect and influences tasks in other phases.

Figure 2 shows the planned thesis approach, divided into three main phases. Each phase tackles a specific part of the thesis, consists of performed activities and influences or is influenced by other activities. In the following, we will explain each phase in more detail.

Phase I: We assume that semantics provide a huge potential for similarity analysis and in revealing the effects on service qualities and process performance indicators. In order to exploit semantics in processes, we first have to annotate current process instances by mapping them to the semantically enriched target model.

The definition of dimensions partially overlaps with the formal specification of processes with semantics. Both activities are performed in phase I.

Phase II: This phase focuses on performing similarity analysis of target process models and current processes. We will use different similarity methods e.g. node similarity, structural similarity and behavioral similarity. Among others, we will also use methods that do not exploit semantics and compare them to methods that exploit semantics in order to show the advantages of having semantic annotations. We will also consider combining different methods for similarity analysis, resulting in a hybrid approach.

Phase III: The last phase uses the similarity analysis as an input in order to reveal the effect on service qualities and dimensions. We will evaluate whether current processes have an influence on the service qualities and dimensions. In addition, during this phase, we can also discover new insights that motivate to capture additional service qualities and dimensions. Therefore, this activity influences in turn phase II.

We will show that the methods are not constrained to a single domain by applying them to different domains (Sect. 6).

5 Preliminary Results

Currently, we are facing the first phase (see Sect. 4), which is about formally specifying process models with semantics. To this end, we analyzed different tools that allow to model processes. However, existing tools do not allow to enrich process data with semantics. In addition, we aim to follow the Linked Data Principles[6] for publishing data.

In order to combine processes with semantic information, we created a tool that captures processes and allow users to enrich them with semantic information. We used bpmn-io as web modeler and extended it with further functionalities. bpmn-io[7] is a JavaScript renderer that allows modeling and checking the syntax of BPMN processes. We embedded our developed tool into a Semantic MediaWiki[8]. Thus, Semantic MediaWiki, in combination with our developed tool, serves as platform to capture, annotate, query and process the information in a structured way and publishing it as Linked Data.

With this tool, we can integrate processes, stored in the standard format BPMN 2.0 XML[9] into Semantic MediaWiki and enrich them with semantics. The integrated and semantically enriched processes can in turn be exported into BPMN 2.0 XML format, allowing for exchange and reuse of the modeled processes.

As the next step, we will determine service quality measures and dimensions for comparing processes but also measuring the efficiency of a target process such as runtime, outcome or costs and study approaches to map current process instances to the target process model. In addition, we will consider different similarity methods to quantify the similarity between target process and current processes and necessary information that will improve the calculation of similarity. These considerations will influence the enrichment of semantic information, since we have to capture it during the annotation of the processes.

[6] http://www.w3.org/DesignIssues/LinkedData.html.

[7] https://github.com/bpmn-io/bpmn-js.

[8] https://semantic-mediawiki.org.

[9] http://www.omg.org/spec/BPMN/2.0/.

6 Evaluation Plan

For validating our solution, we will implement the designed approach and methods in different use-case scenarios. This ensures on the one hand that our approach and methods abstract from the used domain and on the other hand to capture independent results that can be evaluated.

We plan to use the following two domains to evaluate our approach:

(1) Medical Domain: Current processes in clinics differ from target process models. This is caused by latest insights and developments in the medical domain and the slow adoption of clinical pathways. In addition, there are many ontologies i.e. Foundational Model of Anatomy ontology (FMA)[10] or Gene Ontology[11] that can be used to structure processes with semantic information. Therefore, we will use our approach to calculate the similarity between target and current processes and show the influences of processes on different service qualities and dimensions.

(2) Internet of Things: Another field of application is the domain Internet of Things. In this domain, the communication and data flow between devices is not strictly given. Hence, there are more ad-hoc processes, which makes it hard to get an overview of the processes in general. Although this domain is rather new, there are already some ontologies available [32,33]. We will use data from devices (i.e. communication data and process data) and annotate the tasks with semantic information. This allows for enhanced analysis of communication workflows, and allows us to see the deviation of current processes from target process models.

For evaluating the first research question, we will validate the formalized process data, enriched with semantics, by comparing the usability of the provided methods with different approaches and the expressiveness of the formally specified processes. The formalization of data should not be focused on a specific scenario or domain, which is shown by applying our approach and methods in multiple scenarios and domains. Mapping the current process instances to its target process model is validated by comparing it to different methods.

To evaluate the second and third research question, we will start with comparing very simple target and current process models and gradually extend the process with further details and expressiveness. Hence, we will start performing similarity analysis and revealing the effect on different service qualities and dimensions in each applied domain with a sequential process and then successively extend the expressiveness of the process and the used service qualities and dimensions.

7 Conclusions

We aim to develop an approach to annotate and perform similarity analysis between target and current processes.

[10] http://sig.biostr.washington.edu/projects/fm/.

[11] http://geneontology.org.

We will consider the similarity in relation to service qualities and dimensions in order to (1) provide confidence intervals for service qualities and dimensions so one can estimate which values will be assigned by a process variable of the target process model (2) reveal weak spots, which has influence on different service qualities and dimensions and (3) motivate to adapt the target process model if its current process instances diverge too much from it.

In addition, the knowledge from this approach can also be used to support people with process optimization and improvements of the target process models.

References

1. Panella, M., Marchisio, S., Di Stanislao, F.: Reducing clinical variations with clinical pathways: do pathways work? Int. J. Qual. Health Care **15**(6), 509–521 (2003)
2. Kinsman, L., Rotter, T., James, E., Snow, P., Willis, J.: What is a clinical pathway? Development of a definition to inform the debate. BMC Medicine **8**, 31 (2010)
3. Zand, E.K.: Integrated care pathways: eleven international trends. Int. J. Care Coord. **6**(3), 101–107 (2002)
4. Vanhaecht, K., Bollmann, M., Bower, K., Gallagher, C., Gardini, A., Guezo, J., Jansen, U., Massoud, R., Moody, K., Sermeus, W., Zelm, R., Whittle, C., Yazbeck, A.M., Zander, K., Panella, M.: Prevalence and use of clinical pathways in 23 countries - an international survey by the European pathway association. Int. J. Care Coord. **10**(1), 28–34 (2006)
5. Hindle, D., Yazbeck, A.: Clinical pathways in 17 European union countries: a purposive survey. Aust. Health Rev. **29**(1), 94–104 (2005)
6. European Committee for Standardization, Quality management systems - Fundamentals and vocabulary (ISO 9000:2015), September 2015
7. de Nicola, A., Lezoche, M., Missikoff, M.: An ontological approach to business process modeling. In: 3rd Indian International Conference on Artificial Intelligence 2007, pp. 1794–1813, December 2007
8. Smith, F., De Sanctis, D., Proietti, M.: A platform for managing business process knowledge bases via logic programming. In: CILC 2013, pp. 247–251 (2013)
9. van der Aalst, W.M.P.: Process mining in the large: a tutorial. In: Zimányi, E. (ed.) eBISS 2013. LNBIP, vol. 172, pp. 33–76. Springer, Heidelberg (2014)
10. Di Francescomarino, C., Ghidini, C., Rospocher, M., Serafini, L., Tonella, P.: Reasoning on semantically annotated processes. In: ICSOC 2008, pp. 132–146 (2008)
11. Rozinat, A., van der Aalst, W.M.P.: Conformance checking of processes based on monitoring real behavior. Inf. Syst. **33**(1), 64–95 (2008)
12. van der Aalst, W.M.P., Adriansyah, A., van Dongen, B.F.: Replaying history on process models for conformance checking and performance analysis. Wiley Interdisc. Rev. Data Min. Knowl. Disc. **2**(2), 182–192 (2012)
13. Lin, Y., Ding, H.: Ontology-based semantic annotation for semantic interoperability of process models. In: Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, vol. 1, pp. 162–167, November 2005. doi:10.1109/CIMCA.2005.1631259
14. Lin, Y., Strasunskas, D.: Ontology-based semantic annotation of process templates for reuse. In: 10th International Workshop on Exploring Modeling Methods in System Analysis and Design (EMMSAD 2005), Porto, Portugal (2005)

15. Dimitrov, M., Simov, A., Stein, S., Konstantinov, M.: A BPMO based semantic business process modelling environment, semantic business process and product lifecycle management. In: Proceedings of the Workshop SBPM 2007, Innsbruck, April 2007
16. Hepp, M., Roman, D.: An ontology framework for semantic business process management. In: Wirtschatsinformatik Proceedings (2007)
17. European Committee for Standardization, Industrial automation systems and integration – Process specification language (ISO 18629-1:2004), November 2004
18. Bauer, H.H., Falk, T., Hammerschmidt, M.: eTransQual: a transaction process-based approach for capturing service quality in online shopping. J. Bus. Res. **59**(7), 866–875 (2006). ISSN: 0148-2963
19. Gawyar, E.T.H., Ehsani, M., Kozehchian, H.: Measuring service quality of state-clubs in Lorestan Province using SERVQUAL model. Int. J. Sport Stud. **4**(2), 233–237 (2014). ISSN: 2251-7502
20. Collier, J.E., Bienenstock, C.C.: A conceptual framework for measuring e-service quality. In: Proceedings of the Academy of Marketing Science (AMS) Annual Conference, pp. 158–162 (2003). ISSN: 2363-6165
21. del-Río-Ortega, A., Resinas, M., Ruiz-Cortés, A.: Defining process performance indicators: an ontological approach. In: Meersman, R., Dillon, T.S., Herrero, P. (eds.) OTM 2010. LNCS, vol. 6426, pp. 555–572. Springer, Heidelberg (2010). ISSN: 0302-9743
22. Weidlich, M., Sheetrit, E., Branco, M.C., Gal, A.: Matching business process models using positional passage-based language models. In: Ng, W., Storey, V.C., Trujillo, J.C. (eds.) ER 2013. LNCS, vol. 8217, pp. 130–137. Springer, Heidelberg (2013)
23. Dijkman, R., Dumas, M., García-Bañuelos, L.: Graph matching algorithms for business process model similarity search. In: Dayal, U., Eder, J., Koehler, J., Reijers, H.A. (eds.) BPM 2009. LNCS, vol. 5701, pp. 48–63. Springer, Heidelberg (2009)
24. Dijkman, R., Dumas, M., van Dongen, B., Käärik, R., Mendling, J.: Similarity of business process models: metrics and evaluation. Inf. Syst. **36**(2), 498–516 (2011). ISSN: 0306-4379
25. Zhang, Y., Liu, J., Wang, L.: Product manufacturing process similarity measure based on attributed graph matching. In: 3rd International Conference on Mechatronics, Robotics and Automation (ICMRA 2015), June 2015
26. Jung, J., Bae, J.: Workflow clustering method based on process similarity. In: Gavrilova, M.L., Gervasi, O., Kumar, V., Tan, C.J.K., Taniar, D., Laganá, A., Mun, Y., Choo, H. (eds.) ICCSA 2006. LNCS, vol. 3981, pp. 379–389. Springer, Heidelberg (2006)
27. van der Aalst, W.M.P., Reijers, H.A., Weijters, A.J.M.M., van Dongen, B.F., Alves de Medeiros, A.K., Song, M., Verbeek, H.M.W.: Business process mining: an industrial application. Inf. Syst. **32**(5), 713–732 (2007)
28. van Dongen, B.F., de Medeiros, A.K.A., Verbeek, H.M.W., Weijters, A.J.M.M., van der Aalst, W.M.P.: The ProM framework: a new era in process mining tool support. In: Ciardo, G., Darondeau, P. (eds.) ICATPN 2005. LNCS, vol. 3536, pp. 444–454. Springer, Heidelberg (2005)
29. Maggi, F.M., Mooij, A.J., van der Aalst, W.M.P.: Analyzing vessel behavior using process mining. In: van de Laar, P., Tretmans, J., Borth, M. (eds.) Situation Awareness with Systems of Systems, pp. 133–148. Springer, New York (2013)

30. Billington, J., Christensen, S., van Hee, K.M., Kindler, E., Kummer, O., Petrucci, L., Post, R., Stehno, C., Weber, M.: The Petri Net markup language: concepts, technology, and tools. In: van der Aalst, W.M.P., Best, E. (eds.) ICATPN 2003. LNCS, vol. 2679, pp. 483–505. Springer, Heidelberg (2003)
31. Muijs, D.: Doing Quantitative Research in Education with SPSS, 2nd edn. SAGE Publications, London (2010)
32. Hachem, S., Teixeira, T., Issarny, V.: Ontologies for the Internet of Things. In: ACM/IFIP/USENIX 12th International Middleware Conference, Lisbon, Portugal, December 2011. Springer (2011)
33. Wang, W., De, S., Toenjes, R., Reetz, E., Moessner, K.: A comprehensive ontology for knowledge representation in the Internet of Things. In: 2012 IEEE 11th International Conference on Trust, Security, Privacy in Computing, Communications (TrustCom), pp. 1793–1798 (2012). doi:10.1109/TrustCom.20