



Domain Ontology Construction with Activity Logs and Sensors Data – Case Study of Smart Home Activities

Ramona Elali, Elena Kornyshova, Rebecca Deneckere, Camille Salinesi

► To cite this version:

Ramona Elali, Elena Kornyshova, Rebecca Deneckere, Camille Salinesi. Domain Ontology Construction with Activity Logs and Sensors Data – Case Study of Smart Home Activities. 2nd International Workshop on Information Systems Engineering for Smarter Life (ISESL) 2022, CAiSE Workshop, Jun 2022, Leuven, Belgium. hal-03663353

HAL Id: hal-03663353

<https://hal.science/hal-03663353>

Submitted on 23 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Domain Ontology Construction with Activity Logs and Sensors Data – Case Study of Smart Home Activities

Ramona Elali¹, Elena Kornyshova², Rébecca Deneckère¹, Camille Salinesi¹

¹ Paris 1 Panthéon Sorbonne, Paris, France

² Conservatoire National des Arts et Métiers, Paris, France

ramona.elali@univ-paris1.fr, elena.kornyshova@cnam.fr,
rebecca.deneckere@univ-paris1.fr, camille.salinesi@univ-
paris1.fr

Abstract. Process mining relies on activity logs to discover process models, check their conformance, enhance processes, and recommend the next activity. On another side, many environmental factors such as time, location, weather, and profile are obtained from many sources, such as sensors, external systems, outside actors, or domain knowledge bases, and could also enhance recommendations. The existing research mainly focuses on single activity log datasets; only a few consider combining various sources. Our main goal is to provide better inputs to process discovery and better recommendations. In this paper, we focus on the combination of activity logs and sensors data with domain ontology as an intermediate step to attaining our goal. We use a case study of smart home activities to test this combination.

Keywords: Process Mining, Ontology, Sensors, Events Logs, Smart Home.

1 Introduction

An enormous amount of events data and sensor data is being recorded daily whilst the propagation of Information and Communication Technologies (ICT) is taking its rise worldwide across all domains. Data has become available everywhere, hence, the challenge becomes to turn out this data into value to get insight, increase productivity, produce better performance, and save costs [1].

Process Mining (PM) focuses on the generated activities from business processes that can be used as a recommendation technique to direct the user on which next activity to follow according to his current activity [2,3]. PM 's main objective is to discover, check the conformance, and enhance the process models that are based on event logs [1].

Ontology is an important subject in the domain of IS [4]. Many researchers are using ontology to classify domain knowledge [5] such as concepts, all types of entities, and the existing relationships between them [6]. Ontology organizes information in a natural, logical, and systematic way. Therefore, contextual information can be structured into a domain ontology by describing the relationships between the entities of the domain ontology and the contextual information.

In this paper, we propose an approach that combines events logs with contextual data obtained from sensors by using domain ontologies. Nowadays, contextual data can be obtained through many techniques and from various sources. We believe that contextual data can enrich the event logs by providing insights about users' behavior. We test and illustrate this proposal by applying it to a case study of Smart Home daily living activities. We choose this specific case study because this domain is rich in contextual information due to the presence of many sensors. Hence, we use the dataset provided by BP-Meets-IoT 2021 Challenge [7]. This dataset contains an event log of the activities of daily living in addition to the sensor's log data. We are focusing on this case study to construct a specific domain ontology that identifies the links between the different person's activities and the sensor's data.

The next section will focus on the related works. Section 3 will present our proposal which we illustrate in Section 4. We conclude in Section 5.

2 Related Works

PM purposes to enhance, improve, discover, or check the conformance of activity-oriented process models from event logs [1,2]. Event logs are the main input of information for the mining techniques [25].

However, PM approaches do not usually take into account the context aspect behind the recorded user activities. By using various sources of logs, the discovered process models can offer better guidance through processes. In [26], a process mining approach is suggested in the context of sensor data where sensor data were transformed into event logs to be used as input in the PM techniques. A comprehensive review on event logs preparation techniques from raw of data was introduced in [27]. In this review, the authors describe three steps of data preparation: event data extraction from heterogeneous data sources, event correlation, and event abstraction.

Ontology is now widely used in the Information Systems and Artificial intelligence world [6], where many researchers classify concepts, entities, and the relationship between different concepts in domain knowledge [5,6]. An ontology-based personalized recommender approach to provide a more precise and personalized recommendation was presented in [8]. In [9], a combination between the ontology-based recommender system and machine learning techniques is used to provide a personalized recommendation for students. An ontology-based recommendation approach relying on collaborative filtering that achieves a better recommendation in e-Learning systems was presented in [10]. In [11], a classification for research papers was done using ontology then they used collaborative techniques to recommend to the users interesting papers. In addition, activity recognition in Smart Home is a crucial topic to study while working on process mining. In [4], the authors proposed a hierarchical ontology to annotate the human activity in a home environment for healthcare taking into consideration the social context of the activity, the room, and the physical state. [12] developed a knowledge-driven approach by creating ontologies for the sensors and the activities for the purpose of activity recognition. In addition, in [13], they have developed an ontology for the activities of daily living in a smart environment to use as input for a supervised learning technique for online activity recognition. Also, an Activity of a Daily Living ontology-based approach

was developed in [14,15] for activity recognition in a context-aware smart home environment.

3 Approach to Combine Sensors Data and Event Logs Using Domain Ontology

In our view, we believe that ontologies can help to combine sensors data and activity logs. Firstly, all the data acquired from the sensors and all the activities of the events log can be semantically interpreted in an ontology specific to the domain. Secondly, all the links between the ontology concepts (especially the links between the sensors elements and the activities elements) must be defined. For instance, a specific value of sensor data can trigger the start, or the end, of an activity.

To combine sensors data and event logs using domain ontology we have to proceed in three steps (see Fig. 1): (1) create the sensor ontology based on the sensors data, (2) create the activity ontology based on the events log and (3) refine the domain ontology by creating the links between the two preceding ontologies (white arrow on the second part of Fig. 1).

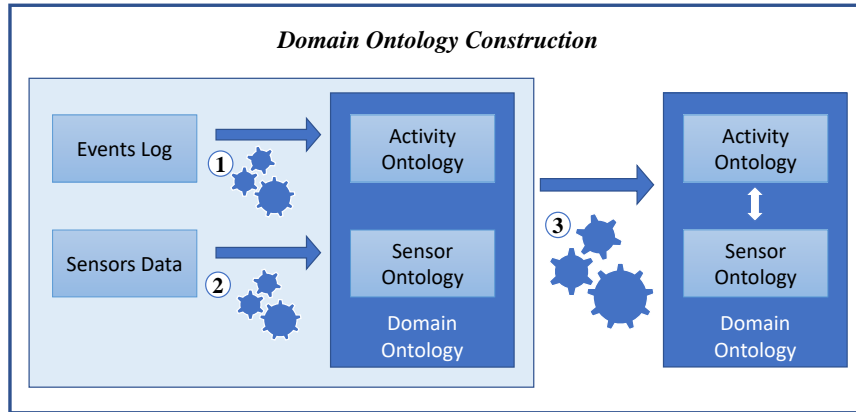


Fig. 1. Domain Ontology Construction Approach.

If the events log and the sensor data are related to existing ontologies, we can proceed directly to step 3 to identify the links between them.

In this work, we focus on step 3 with two main hypotheses: **H1:** *It is possible to identify the ontological links between sensors data and activities using the timestamp on data and events.* **H2:** *It is possible to identify the ontological links between sensors data and activities by association rules.*

Timestamps. The timestamp is considered the only attribute that can let us find the relationship between the activity event and the sensor. We developed a Python algorithm that takes as input the events log and sensor log. The two logs need to be

sorted according to the timestamp field. This algorithm is composed of two main steps. In the first step, the start and completion of each activity should be identified. Then, the timestamp field is used to pivot the events log according to the start and completion of each activity. Then in the second step, a join was done between the two logs using the timestamp field to create the link between the sensor data and the event data.

Association rules. Association rule mining is the technique of finding relationships between different items. It is commonly used in Market Basket Analysis, where it identifies which items are frequently purchased together for marketing purposes. It is the process of finding frequent patterns, associations, causal structures, and correlations within different datasets existing in different types of databases like transactional databases, relational databases, or any other form of databases [16, 17]. The purpose of association rule mining is to find the rules that allow us to guess the appearance of a specific item after the occurrences of other items in the transaction. It is one of the top concepts of machine learning rule-based techniques that discover interesting relations between items in big databases by indicating the frequent associations' rules that are composed of an antecedent and consequent [16]. Association rules allow uncovering the hidden relationships between the items that are frequently used together. It is represented by if-then statements (an antecedent itemset and a consequent itemset). Apriori [18] is one of the top standard algorithms for frequent itemset mining and association rule learning. In this work, we used Association Rule mining techniques since we needed to find rules between the sensors and the activities. To apply the Apriori algorithm, first of all, we needed to transform the dataset into a transaction log that is supported by the algorithm. We labeled each log by day date and then we combined both logs and sorted them according to the day and the timestamp. Then we transformed the combined log into a transaction log which is supported by Apriori. In addition to the transaction log, Apriori needs as input multiple parameters such as the minimum support and minimum confidence. *Support* indicates how often an itemset is frequent in the transactions. *Confidence* indicates how often the rule (if \rightarrow then) statement is found true.

We illustrate our proposal through a Smart Home case study in the following section.

4 Case Study

The rapid evolution of artificial intelligence endorses the existence of smart homes and make it crucial to study this domain. Numerous studies are done on smart home [19,20,21,22]. Hence, our case study will be focused on smart home since this field is rich in different types of sensors. We need to find the link between the person's activities and the sensors to find later how the context can affect the user's intention. Data preprocessing includes standard activities aiming at enhancing data quality: cleaning, integration, and transformation [23].

In our experiment, we are using the dataset provided by BP-Meets-IoT 2021 Challenge [7]. This dataset consists of two people's activities logs in addition to the sensor logs recorded during 21 consecutive days between 16 March 2020 and 6 April 2020. While

studying the dataset, we found 53 activities and 16 sensors. Both logs contained noise data; some duplicated records, in addition to some records that don't represent any type of activities or sensors. We cleaned the dataset from the noise by removing those tuples, and we obtained 23166 records in the events log and 39304 records in the sensors log. A sample of each of the datasets is shown in Fig.2. Each record in the event logs is identified by the Activity Name while each record in the sensor logs is identified by the Sensor Name and both logs are categorized by the Timestamp and the Resource Id.

Event Logs:

ActivityName	ResourceId	Timestamp	Transition	EventId	CaseName	TraceId	Position	Day
go_wardrobe	2	2020-03-16 00:00:00+00:00	complete	4071	sleeping_BP	435	wardrobe	Day1
get_clothes	2	2020-03-16 00:00:00+00:00	start	4072	sleeping_BP	435	wardrobe	Day1
get_clothes	2	2020-03-16 00:01:00+00:00	complete	4076	sleeping_BP	435	wardrobe	Day1
change_clothes	2	2020-03-16 00:01:00+00:00	start	4077	sleeping_BP	435	wardrobe	Day1
change_clothes	2	2020-03-16 00:05:00+00:00	complete	4080	sleeping_BP	435	wardrobe	Day1
go_bathub	2	2020-03-16 00:05:00+00:00	complete	4081	sleeping_BP	435	bathroom_sink	Day1

Sensor Logs:

SensorName	ResourceId	Timestamp	EventId	Day	Value	Position
position	1	2020-03-16 00:01:00+00:00	18	Day1	1202 852	wardrobe
position	1	2020-03-16 00:06:00+00:00	24	Day1	1184.75270036629 848.333251258976	
water_use	0	2020-03-16 00:07:00+00:00	25	Day1		10
position	1	2020-03-16 00:07:00+00:00	27	Day1	1246 638	wc
water_use	0	2020-03-16 00:31:00+00:00	51	Day1		1
position	1	2020-03-16 00:31:00+00:00	53	Day1	1164 768	bathroom_sink

Fig. 2. Event Logs and Sensor Logs Samples.

There was not any defined domain ontology given with the dataset. Moreover, despite previous works already having defined ontologies for the daily activity recognition in smart homes [4,12,13,14,15], none of them were matching the data of our dataset. Thus, we had to define this domain ontology ourselves. After a deep analysis of the data of the sensors and events log, we were able to identify the different parts of the domain ontology. To construct our domain ontology - which is specific to our dataset and case study - we have defined two main ontologies (Smart home activity ontology – sub-section 4.1 and smart home sensor ontology – sub-section 4.2). Their links are detailed in sub-section 4.3. To ease the understanding of our work, we propose here only a simplified view of the two ontologies where all the links between the concepts are “is-a” relationships. We define the more complex associations between the concepts in a specific meta-model (cf. Fig. 10).

4.1 Smart Home Activity Ontology

The Activity ontology (Fig. 3) contains the different types of activities and actions that can happen inside a home.

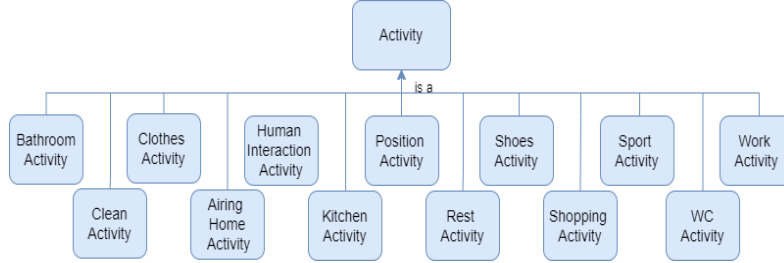


Fig. 3. Smart Home Activity Ontology.

Each activity should be completed by a human inside a fixed location. The Activity ontology is divided into groups of activities (see Fig. 3): Bathroom Activity, Clean Activity, Work Activity, etc... Each activity group inside the Activity ontology has its own instances. As an example, the Kitchen Activity has a set of real-life activities that happen inside the kitchen (see Fig. 4). Drink Water, Go Kitchen Shelf, Put Plate to Sink, Wash Dishes, etc... are instances of Kitchen Activity group.

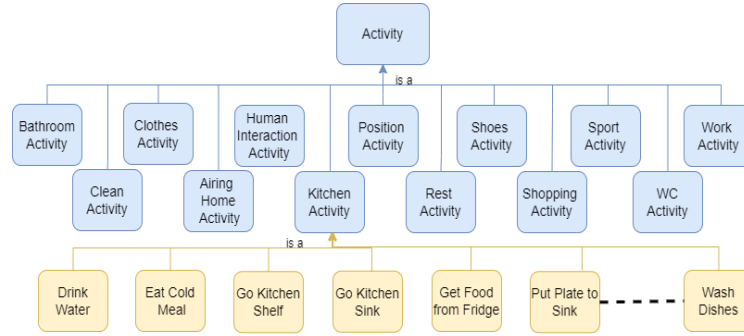


Fig. 4. Smart Home Activity Ontology with Details about the Kitchen Activity.

4.2 Smart Home Sensors Ontology

Smart Home Sensors Ontology includes three parts: Sensor, Position, and Human ontologies.

The sensor ontology contains all the information related to the sensors such as sensor type, value, value range, and location. According to our dataset, we were able to identify different types of sensors such as Environment sensors, Kitchen Sensors, etc. In Fig. 5, we have presented the high level of the ontology sensor, where the sensors are divided into different types and each type contains its related sensors.

For each location inside the home, there are a set of sensors related to the location such as Kitchen Sensor, Bedroom Sensor, and Living Room Sensor. And there are some sensors that are common to all the house as the Environment Sensor, Power Sensor, and Water Sensor.

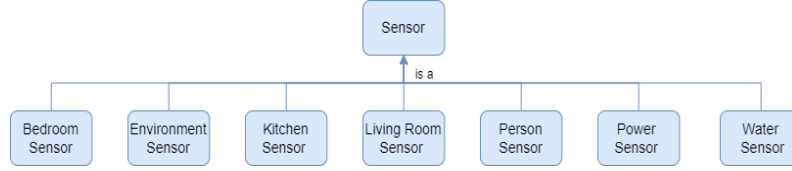


Fig. 5. Sensor Ontology High Level.

Fig. 6 shows an example of instances of Kitchen Sensor. Fridge Door Contact Sensor, Cooked Food Sensor, Food Sensor, and Unwashed Dishes Sensor are all sensors located inside the kitchen. Each sensor works differently according to its type, and it has its own range of values [7]. For instance, the Unwashed Dishes sensor is used to count the number of unwashed dishes inside the kitchen sink, its value is greater than 0 when there are dishes inside the sink to wash, otherwise, it will be 0. While the Fridge Door Contact Sensor indicates when the fridge door is opened or closed, so its value is 0 when it is closed and 1 once it is opened. The Food Sensor is used as a food counter inside the home, so its value is equal to or greater than 0. As for the Cooked Food Sensor, it works the same as the Food Sensor, but it counts only the available cooked food inside the home.

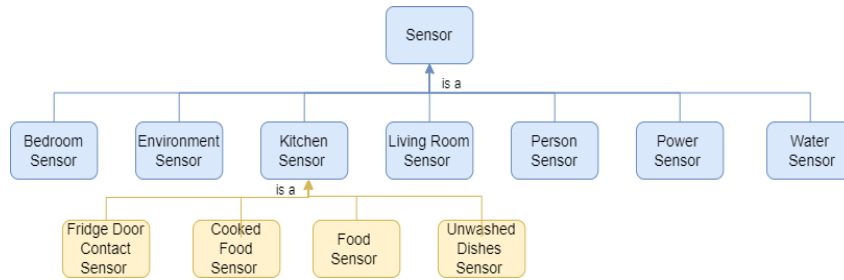


Fig. 6. Kitchen Sensor Instances Example.

Position ontology (Fig. 7) represents the different parts of the home where the activities can take place and the sensors are attached. Position ontology is divided into 6 main positions categories (see Fig.8): Entrance Position, Exercise Place Position, Bathroom Position, Kitchen Position, and Living Room Position.

Each position category has its own instances that represent the different fixed locations inside the home. As an illustration, the Kitchen Position contains the different positions inside the conditions such as the dining table, the fridge, the kitchen shelf, and the kitchen sink (see Fig. 8).

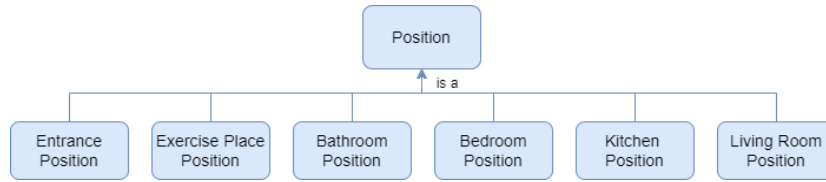


Fig. 7. Position Ontology High Level.

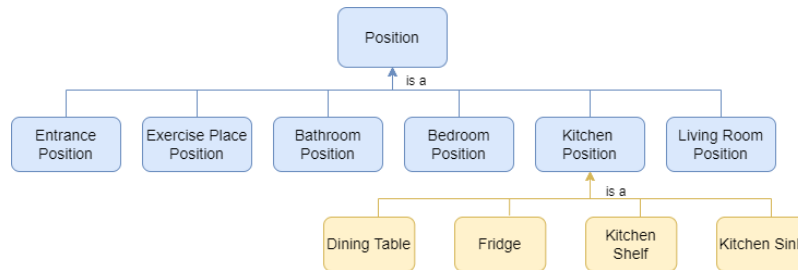


Fig. 8. Kitchen Position Instances Example.

Then, Human ontology is a needed domain to have to characterize the human profile (see Fig. 9). It categorizes the different person types that perform the different activities. It can be divided into two main subdomains: the initial home residents (inhabitants) and the guests (see Fig. 9).

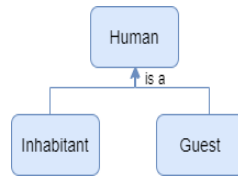


Fig. 9. Human Ontology High Level.

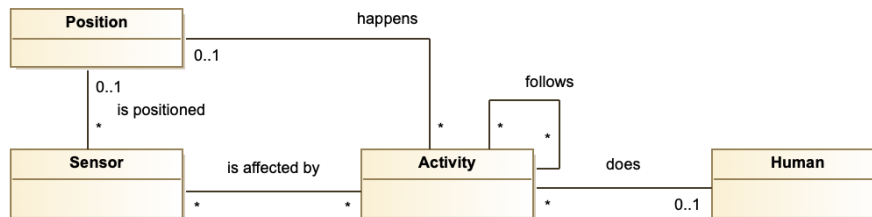


Fig. 10. Links between the Ontologies.

4.3 Links between Sensors and Activity Ontologies

We can identify different relationships between the defined ontologies (see Fig. 10). Each human does activities. Sensors affect activities and can be positioned at a specific position. Each activity can be followed by one or many activities (that can be done sequentially or not). Each activity can happen at a particular position at home and be affected by sensors. Each position can be related to sensors and activities.

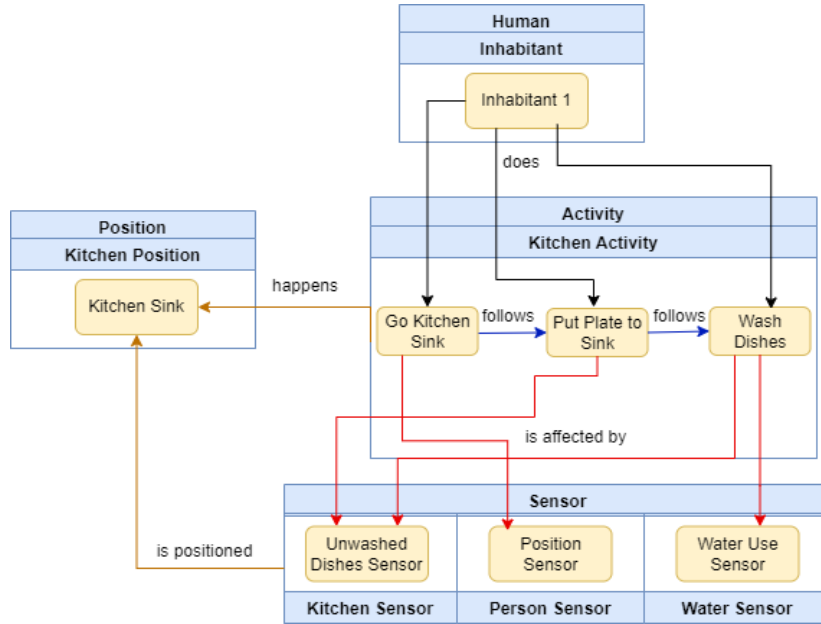


Fig. 11. Example on the Link between the Different Ontologies' Domains.

Fig. 11 illustrates an example of the links between the ontologies. When the human 'goal is Washing Dishes then the set of activities that are related to this goal will take place inside the kitchen; the inhabitant will go to the kitchen sink, put the plate inside the sink and start to wash dishes. In the meantime, the sensor of Unwashed Dishes will increase its value according to the plates that are being washed. The Water Use sensor will increase its value and the Position sensor will indicate that the human position is next to the Kitchen Sink.

The established relationships were used to test the established hypotheses. As presented in Section 3, we had two main hypotheses: it is possible to identify the ontological links between sensors data and activities using the timestamp on data and events (H1), and it is possible to identify the links by association rules (H2).

To validate H1, an automatic test was done using Python as a programming language. 100836 records were obtained corresponding to 53 different activities and 16 different sensors. The obtained data was cleaned because it was containing some noise data thus the remaining was 99975 records. We obtained 112 correspondences between

activities and sensors. We compared the mapping results with the results that we obtained manually while analyzing the data and deduced that for each occurred activity there is one or multiple sensors triggered. An extraction is given in Table 1 to illustrate the obtained links between activities and sensors using the timestamp data (the complete list can be found in [24]). In this manner, we were able to validate H1 and identify the link between the activities and the sensors while relying only on the timestamp.

Table 1. Sample of Activity and Sensor Mapping.

Activity	Sensor
Brush Teeth	Position Sensor Water Use Sensor
Change Clothes	Position Sensor Pressure Bed Sensor
Clean	Position Sensor Home Aired Sensor
Close or Open Windows	Windows Sensor Blinds Sensor
Washes Dishes	Unwashed Dishes Sensor Position Sensor Water Use Sensor
Eat Cold Meal	Food Sensor
Get Food from Fridge	Fridge Door Contact Sensor Position Sensor

To validate H2, a second test was done using Association Rules by applying the Apriori algorithm [18] on the dataset selecting only the logs that are relevant to one person. After combining the logs, we obtained 410 transaction records. We did three experiments. We set the confidence value to 80% in all the experiments to get strong rules. The minimum support value is set through the process of trial and error. In the first experiment, we set the value of minimum support to 0.15 and we obtained 56 rules. In the second experiment, we set the minimum support value to 0.05 and we obtained 8565 rules. And in the third experiment, we set a small value of minimum support equal to 0.03 and we obtained 34623 rules. We analyzed the generated rules manually and found that the result of the second experiment is more realistic due to the value of the minimum support and the confidence. In Table 2, we present a small extraction of the obtained rules [24] with the corresponding confidence degrees for each rule.

Thus, despite the limitation of Apriori to perform well when the itemset is large, we were able to generate rules that associate activities with sensors and validate H2.

On the whole, this experiment has validated the existing relationship between the sensors and the activities. We were able to automatically find the rules that we have found before manually while analyzing the dataset. For an instance, when a person goes to the kitchen next to the shelf; the position sensor will locate his position accordingly, so the next activity will be to get a glass then move to the sink where the ‘water use’ sensor will directly reflect the water consumption while filling the glass with water. And after drinking water, the ‘unwashed dishes’ sensor will increment its value.

Table 2. Sample of the obtained rules.

Rule	Confidence
{eat_cold_meal_go_kitchen_shelf, go_fridge, fridge_door_contact_sensor} → {position_fridge_sensor, food_sensor, get_food_from_fridge}	0.92
{position_kitchen_sink_sensor, water_use_sensor} → {drink_water }	0.92
{go_computer_chair, go_computer, position_computer_chair_sensor }→ { switch_computer_on, power_use_sensor}	0.83

5 Conclusion

This research paper describes an approach to constructing domain ontologies using sensors data and events logs. Sensor, Position, Human, and Activity are the ontologies that were defined. Our main objective in this work was to find relations between sensors and activities. We were able to identify and validate the link between sensor data and activities, based on a Smart Home case study. However, the limited amount of data prevented us to find more interesting links between activities and sensors.

A future experiment on a larger dataset will provide us with a more complete set of links. Then, we will continue our main research proposal by using these ontologies as input to build process models by precisising the links between sensors and activities in order to identify which data can trigger an activity. Our main goal is to be able to make recommendations to users on the fly, following these contextual data and the event logs.

References

1. Aalst, W.M.P.: Process Mining: Data Science in Action. Springer, Heidelberg ISBN: 978-3-662-49850-7 (2016).
2. Khodabandelou, G., Hug, C., Deneckère, R., Salinesi, C.: Process Mining Versus Intention Mining. Lecture Notes in Business Information Processing. 147. 466-480 (2013).
3. Compagno, D., Epure, E., Deneckere, R., Salinesi, C.: Exploring Digital Conversation Corpora with Process Mining Corpus Pragmatics, Springer, 2, pp.1-23 (2018)
4. Woznowski, P., King, R., Harwin, W., Craddock, I.: A Human Activity Recognition Framework for Healthcare Applications: Ontology, Labelling Strategies, and Best Practice. 369-377, (2016).
5. Mansingh, G., Osei-Bryson, K. M., Reichgelt, H.: Using ontologies to facilitate post-processing of association rules by domain experts. Information Sciences, Volume 181, Issue 3, Pages 419-434, (2011).
6. Viinikkala, M.: Ontology in Information Systems. In Computer Science (2004).
7. Koschmider, A., Leotta, F., Serral, E., Torres, V.: BP-Meets-IoT 2021 Challenge Dataset (2021).

8. Ge, J., Chen, Z., Peng, J., Li, T.: An ontology-based method for personalized recommendation. 2012 IEEE 11th International Conference on Cognitive Informatics and Cognitive Computing, Kyoto, Japan, pp. 522-526 (2012)
9. Obeid, C., Lahoud, I., Khoury, H., Champin, P.A.: Ontology-based recommender system in higher education. 2018 Web Conference Companion (WWW 2018), France, (2018).
10. Zhang, Z., Gong, L., Xie, J.: Ontology-Based Collaborative Filtering Recommendation Algorithm. 7888. 172-181, (2013).
11. Middleton, S., De Roure, D., Shadbolt, N.: Ontology-Based Recommender Systems. (2009).
12. Chen, Liming & Nugent, Chris & Wang, Hui. A Knowledge-Driven Approach to Activity Recognition in Smart Homes. Knowledge and Data Engineering, IEEE Transactions on. 24. 1 - 1. 10.1109/TKDE.2011.51. (2012).
13. Salguero, A., Espinilla, M., Delatorre, P., Medina, J.: Using Ontologies for the Online Recognition of Activities of Daily Living. Sensors. 18. (2018).
14. Chen, L., Nugent, C.: Ontology-based activity recognition in intelligent pervasive environments. IJWIS. 5. 410-430. (2009).
15. Bae, I.: An ontology-based approach to ADL recognition in smart homes. Future Generation Computer Systems. 33. 32–41. (2014).
16. Zaki, J., M., Meira J., W.: Data Mining and Analysis: Fundamental Concepts and Algorithms (2014).
17. Sunita, D., A., Lobo, L., M., R., J.: A Comparative Study of Association Rule Algorithms for Course Recommender System in E-learning. International Journal of Computer Applications. 39. 48-52. 10.5120/4788-7021. (2012).
18. Agrawal, R., Srikant, R.: "Fast algorithms for mining association rules". In Proc. Int'l Conf. Very Large Data Bases (VLDB), pages 487–499, Sept. (1994).
19. Tiwari, P., Garg, V., Agrawal, R.: Changing World: Smart Homes Review and Future, EAI/Springer Innovations in Communication and Computing, (2022).
20. Alam, M., R., Reaz, M., B., I., Ali, M., A., M.: A review of smart homes - Past, present, and future, IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, (2012).
21. De Silva, L., C., Morikawa, C., Petra, I., M.: State of the art of smart homes, Engineering Applications of Artificial Intelligence, (2012).
22. Alaa, M., Zaidan, A., A., Zaidan, B., B.; Talal, M., Kiah, M., L., M.: A review of smart home applications based on Internet of Things, Journal of Network and Computer Applications, (2017).
23. Tamilselvi, R., Sivasakthi, B., Kavitha, R. An efficient preprocessing and postprocessing techniques in data mining. International Journal of Research in Computer Applications and Robotics, 3(4), 80-85. (2015).
24. Elali R.: Data Exploration of BP-Meets-IoT 2021 Challenge Dataset: Correspondences & Associations Rules Result, White paper, University of Paris 1 Panthéon-Sorbonne (2022).
25. Diaz-Rodriguez, O. E., Perez, M., Lascano, J.: Literature Review about Intention Mining in Information Systems. Journal of Computer Information Systems ,1-10. (2019).
26. Van Eck, M. L., Sidorova, N., Van der Aalst, W. M.: Enabling process mining on sensor data from smart products. In 2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS) (pp. 1-12). IEEE. (2016, June).
27. Diba, K., Batoulis, K., Weidlich, M., Weske, M.: Extraction, correlation, and abstraction of event data for process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1346. (2020).