An ontology-based approach for building automation data analysis

Eunju Park [a], Sumee Park [b], Sebastian Stratbücker [c]

[a] Thermal Comfort, Models and Simulation, Department Energy Efficiency and Indoor Climate, Fraunhofer Institute for Building Physics IBP, Valley, Germany, eunju.park@ibp.fraunhofer.de

[b] Thermal Comfort, Models and Simulation, Department Energy Efficiency and Indoor Climate, Fraunhofer Institute for Building Physics IBP, Valley, Germany, sumee.park@ibp.frauhnhofer.de

[c] Thermal Comfort, Models and Simulation, Department Energy Efficiency and Indoor Climate, Fraunhofer Institute for Building Physics IBP, Valley, Germany, sebastian.stratbuecker@ibp.fraunhofer.de

Abstract. For the efficient and sustainable operation of building automation systems, it is critical to consider various aspects such as users' comfort requirements and energy consumption. The successful application is associated with the integration of multiple and heterogeneous data sources. However, the high complexity of the data poses a challenge. To address this problem, various ontologies have become popular with many applications for data modelling, management and analysis through harmonizing different data sources, as well as efficient querying. In this work, the design, implementation and usage of semantic approaches is investigated to exploit building automation data for customized room automation. As a main contribution, a building automation ontology focusing on room automation is proposed, which is represented in the Resource Description Framework (RDF). Furthermore, several scenarios with the proposed model are demonstrated in-situ, showing easier access to various data sources using a query language like SPARQL. Based on the ontology in RDF format, building data from different sources such as commercial building automation system (e.g. KNX), weather station and room monitoring sensors (e.g. temperature, humidity) are considered for multiple scenarios: (a) anomaly detection of shading automation systems, (b) monitoring user's shading controls in automatic and manual mode, (c) identifying influential factors affecting user's preference. The ontology-based approaches have benefits especially in multiple and heterogeneous data environments using a standardized common and controlled vocabulary. It allows engineers and researchers to enrich and interlink with various databases. Additionally, it explicitly describes the relationships between variables that make data understandable for both humans and machines.

Keywords. Ontology, Semantic model, Anomalies detection, Data mining,
DOI: https://doi.org/10.34641/clima.2022.351

1. Introduction

There is increasing demand for building automation systems. These play an important role in solving existing global warming challenges. Simultaneously, they need to meet user expectations for comfort and usability. The successful application of efficient and sustainable building operation is associated with integration of multiple and heterogeneous data sources. However, the high complexity of the data remains a big challenge. To address these problems, the application of ontologies has become popular in data modelling, management and analysis by harmonizing different data sources, as well as efficient querying. Successful application requires appropriate ontologies to create and accumulate building data through standardized metadata schemas.

Numerous ontologies have been introduced provide the standardized metadata to represent knowledge of buildings with terms in formal and shared conceptualization. As discussed in Pritoni et al. [1], several communities use individual terms, resulting in 40 metadata schemas that adopted different standards. Building Topology Ontology (BOT) [2] provides topological concepts of building components by defining relationships between subcomponents: Site, Building, Storage, Space and Element. The Smart Applications REFerence (SAREF) [3] was intended to capture generic sensors and smart devices, which allows the interoperability among IoT solutions developed by different manufacturers. However, it does not effectively cover the diversity of devices and equipment in buildings. Brick [4] aims to provide a standardized ontology for physical, logical, and virtual assets in buildings and the relationships between them. Frequently utilized, it shows extensibility and interoperability with other existing schemas. Existing schemas are strong candidates to digitalize building data. However, they are limited to describing building concepts, ignoring logics (e.g. rule-based automation).

To propose a building automation ontology focusing on room automation: (a) the existing schemas are reused, (b) a new schema is created to represent motions triggered by automation systems and user's commands, and (c) these schemas are integrated. With the ontology model investigates the design, implementation and usage of semantic approach to exploit building automation data.

The motivation of this work is to convert existing data sources from a relational database to a model based on the Resource Description Framework (RDF). This allows engineers and researchers to engage with multiple databases, enabling further analysis. Once the data in a RDF format is stored in a graph database, it can be easily accessed on the web using a query language (e.g., SPARQL). In order to demonstrate the application of the RDF data model, building data such as commercial building automation system (e.g. KNX), weather station and room monitoring sensors (e.g. temperature, humidity) are considered in several scenarios: (a) detecting anomalies of shading automation system based rules by monitoring whether the automation system works properly, (b) monitoring user's activities in shading controls when it is in automatic and manual mode, (c) identifying influential factors influencing user preference.

The proposed ontology is illustrated in detail and the usefulness of the proposed model is evaluated in case scenarios.

2. Ontology

The proposed ontology aims at combining the following information:

- Topological concepts of a building
- Equipment, devices and their physical location in a building as well as logical relationships between them
- Rule-based logics of automation systems and actual motions
- Data sources

According to the ontology development guide [5], reusing existing ontologies was considered. The scope includes motions triggered by automation systems and users' commands. The prefixes of the reused ontologies are in Tab. 1.

Tab. 1 – Prefixes and namespaces

Prefix	Namespace
brick	https://brickschema.org/schema/Brick#
bot	https :// w3id .org/bot#
owl	http://www.w3.org/2002/07/owl#
qudt	http://qudt.org/schema/qudt/
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
skos	http://www.w3.org/2004/02/skos/core#
tag	https://brickschema.org/schema/BrickTag#
unit	http://qudt.org/vocab/unit/
xsd	http://www.w3.org/2001/XMLSchema#

2.1 Building structure

In the proposed ontology, well-known concepts such as basement, rooftop and floor form the basic building structure. Several spaces belong to each floor to describe building indoor parts. Spaces are specified into subclasses (e.g. Office). For example, a floor in a building has East and South zones and offices in the East zone. The elements are linked to form the hierarchical building structure by using relationships bot:hasStorey, bot:containsZone or bot:hasSpace. Devices and equipment can be assigned directly to basement, rooftop, floor, zone and space by using relationships brick:hasLocation or brick:isLocationOf. Then, the devices and equipment in the same location can be easily inferred.

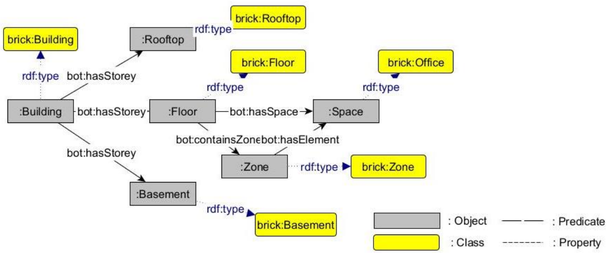

Fig. 1 – Structural modelling of a building.

2.2 Equipment and devices

Equipment and devices are assigned to target locations. Each office has 4 systems: shading, lighting, HVAC and occupancy. Each system except for occupancy consists of 4 components.

- Target object: when an actuator receives a command, it produces a motion to the target object (e.g. blind, light).
- Sensor: to measure the indoor climate. Each sensor belongs to a corresponding system.
- Manual control: input produced by a user manually through a manual object (e.g. switch).
- Remote control: input produced remotely

from a remote device (e.g. PC, mobile devices).

- Motion: actual motion triggered by rules or commands by users. Motion classes are explained in section 2.3 in detail.

Fig. 2 – A room with indoor climate monitoring devices

Each room has a light, blind and radiator as a target object (Fig. 2). The status of target objects are monitored: light (light power), blind (blind position, blind angle), radiator (temperature measured via thermostat, valve position). Sensors to measure indoor climate are located in each room.

- Lighting: Illuminance sensors on desk and ceiling
- HVAC: Temperature, CO2, humidity sensors
- Presence: Occupancy sensors

Commands issued by a user are monitored separately depending on device types (physical switch or remote interface such as PC, app, etc.). A setpoint (e.g. desired radiator temperature) is considered as commands.

- Lighting: Light on/off, dimming
- Shading: Blind up/down, angle position
- HVAC: Desired radiator temperature

A real room in Fig. 2 is described as an ontology as illustrated in Fig. 3.

The weather station located on the rooftop consists of several types of sensors as illustrated in Fig. 4. Temperature, illuminance and wind speed sensors are included in the ontology. Other sensors (e.g. wind direction, air pressure) could increase flexibly. As systems and devices differ, other concepts can be added if additional characteristics are desired.

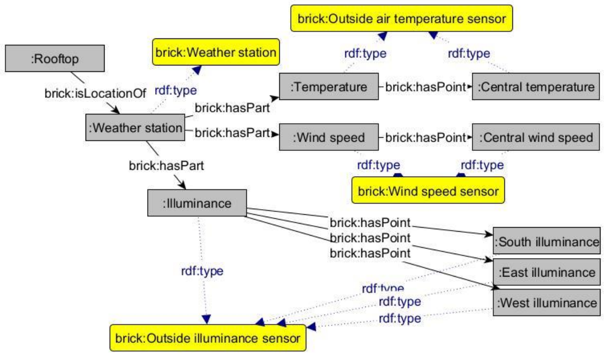

Fig. 4 – Weather station on the rooftop

2.3 Motion

Although there have been several attempts to describe rule-based algorithms in building automation [8], [12], no ontology exists to define motions of an actuator. An ontology is created to express lighting, shading and HVAC system motions. Two factors initiate a trigger to an actuator: rule and command. An object defined as motion class refers to motions triggered by rules of automation systems or commands by users. Similarly to existing ontologies, it has a hierarchical structure. As shown in Fig. 5, the Motion is defined as the top class with subclasses defining specific types of motions: shading, lighting and air control motions.

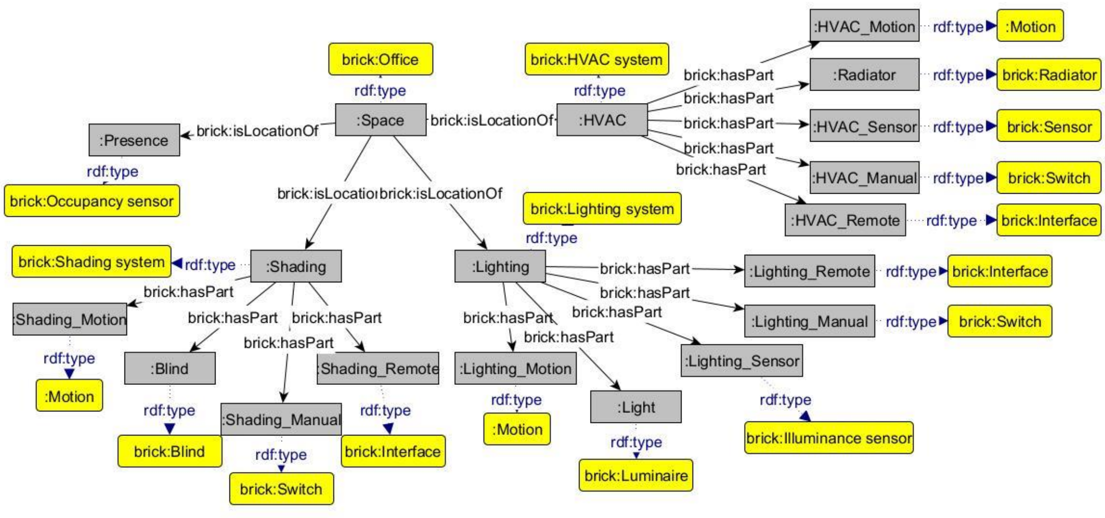

Fig. 3 – Indoor climate monitoring systems

By using the motion ontology and existing ontologies, relationships between motion and rule or command in a room are represented. An example of relationships is presented in section 2.5.

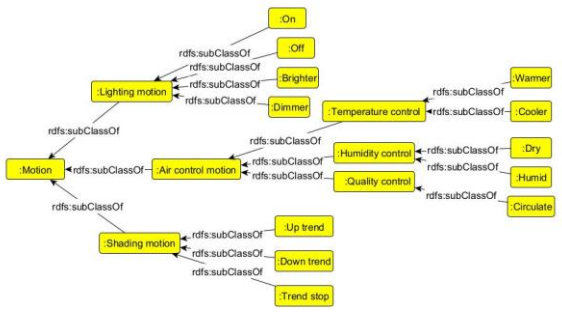

Fig. 5 – Motion ontology

2.4 Data sources

Since the intention was to design an ontology for easier access to relational databases, the data source information is represented in the ontology. As time series data, each data point has at minimum, a timestamp and attribute. If data of an object is available in a relational database, the object includes data source information: database name and primary key. For data description, the unit type information (e.g. Celsius, LUX) is added. For instance as seen in Fig. 7, outdoor illuminance data is collected from a weather station. It is stored in a database called *wetterturm*. The illuminance sensor has the data source and unit (LUX) information as properties.

Although the source information is simplified with database name, primary key and unit type, low-level description such as IP address can be included as well.

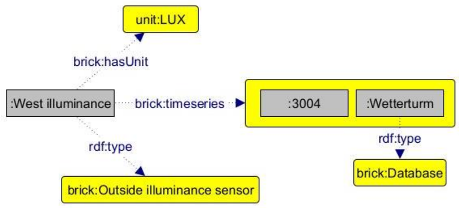

Fig. 7 – Data source information of illuminance sensor

2.5 Example: Shading system

In this section, relationships in up-trend motion in shading systems are described. As shown in Fig. 8, Up-trend of a shading system in a room receives inputs from commands (:Push button up) and an illuminance sensor located on the rooftop. It is configured to trigger the motion depending on the outside illuminance level.

When the conditions are satisfied, the automation system opens the blind. Simultaneously, commands are considered as inputs. A single command input changes the position of the blind. It means that a user can not only trigger a motion by using a manual switch or a remote interface, but also interrupt Up-trend motion triggered by the automation system. As the motion is triggered, it is transferred to the target object (blind). At the same time, the physical position or status of the target object is monitored. If an object has data, it has a property brick:timeseries to refer the database and unique key used in a relational database the database and unique key used in a relational database. The data related to shading system in a room is stored at two different databases: :imedas and :Wetterturm at the investigated building.

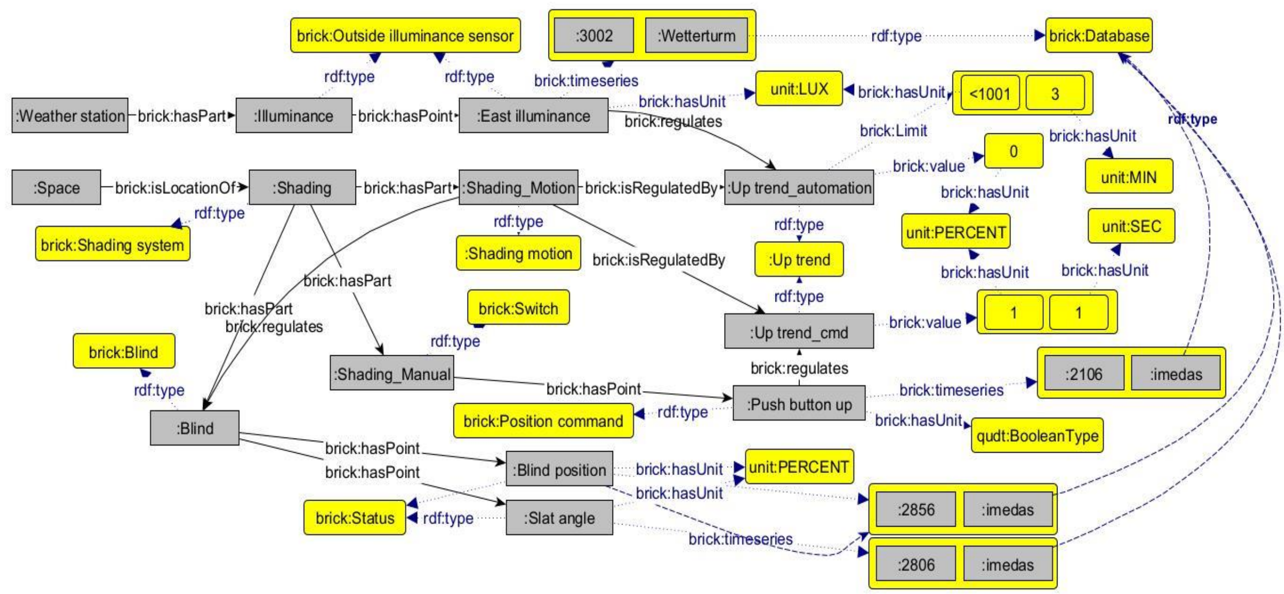

Fig. 8 – Up-trend motion in shading system trigged by automation system (outside illuminance level) or user (manual switch)

3. Application

In the following section, several useful applications are introduced based on the proposed ontology in shading automation systems. Since a user is not regularly present at work, different target offices and periods are used in each application.

3.1 Current building & Shading automation

Twenty offices located on first or second floor in the same building are monitored. Each office has 40 or 41 channels (e.g. sensors, commands) for real-time monitoring. Each office has different numbers of channels monitored. Regardless of the different numbers, the data is easily accessible based on the ontology.

The outside illuminance level and wind speed are only used as the configuration values in shading automation systems. When the outside illuminance level is smaller than 1001 Lux or wind speed is over 10 m/s, the automation system changes the blind position to 0(open). Likewise, it changes the blind position to 100(close) when the outside illuminance level is bigger than 25000 Lux.

3.2 Anomaly detection of shading automation system

From July 2nd to 6th July, two offices (E2, E3) facing the east direction and 2 offices (S2, S3) facing the south direction were monitored. If the configuration condition is satisfied, the position of a blind is changed by the automation system. Since there was no strong wind, only outside illuminance level was influential. Here anomalies are defined as irregular behaviours of shading automation systems, although the configuration conditions are satisfied in automation mode.

The actions triggered by automation systems in office E2 and E3 were identical. Likewise, office S2 and S3 reported identical automation operations. While the Down-trend motion (blind closing) was configured in east-facing offices at 25000 Lux, the blinds were closed in south-facing offices when the outside illuminance level was higher than 35000 Lux. The opening trigger illuminance level was identical with 1000 Lux for both façade.

It took the same time to deliver the operation to south and east-facing offices. Except 5th and 6th July in office E2 and E3 between 5:21 A.M. and 5:25 A.M., approximately 5 minutes delay existed in general. In spite of the delay, it means that the automation system works properly as it is configured.

As the ontology has descriptive information including relevant elements, data source information and configuration values in each office, it is easily retrieved by using a SPARQL as written in Frag. 1 and Frag. 2. Fig. 10 illustrates the filtered sample offices E2 and S2 in shading automation system.

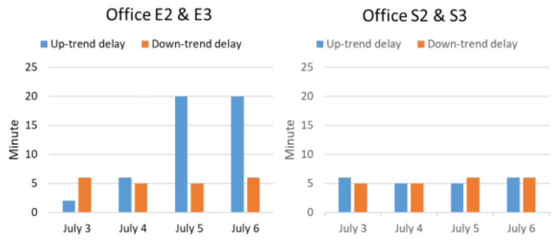

Fig. 9 – Observed command passing delay during 2nd to 6th July.

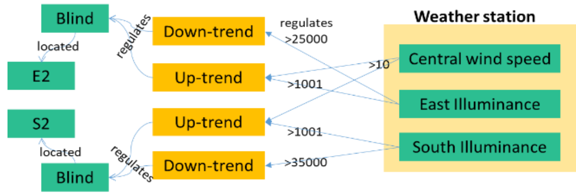

Fig. 10 – Retrieved results of E2 and S2.

Frag. 1 – SPARQL query to get the status of blinds and their data source information.

```
SELECT ?office ?stat ?db ?key
WHERE {
?office rdf:type/rdfs:subClassOf*
brick:Office .
?blind a brick:Blind .
?blind brick:hasLocatoin ?office .
?blind brick:hasPoint ?stat .
?stat a brick:Status .
?stat brick:timeseries ?uid .
?uid brick:hasTimeseriesId ?key .
?uid brick:storedAt ?db . }
```

Frag. 2 – SPARQL query to get configuration parameters and related data source information.

```
SELECT DISTINCT
        ?office ?motion ?point ?db ?key ?val
WHERE {
?sys brick:hasLocation ?office .
?sys brick:hasPart ?elem .
?elem a mo:Shading_Motion .
?elem brick:isRegulatedBy ?motion .
?motion brick:isRegulatedBy ?point .
?point brick:timeseries ?timeseriesid .
? timeseriesid brick:hasTimeseriesId ?key .
? timeseriesid brick:storedAt ?db .
?point brick:hasUnit ?punit .
?motion brick:Limit ?limit .
?limit brick:hasUnit ?lunit .
?limit brick:value ?val .
FILTER (?punit = ?lunit)
}
```

3.3 Shading control in automatic and manual mode

The indoor climate and user's behaviours were monitored during July (10 days). In this section, two users from east-facing office E1 and south-facing office S1 and their offices are described in shading systems. While the shading automation in E1 was active and the position of the blind was controlled by the automation system as it is configured during the observed period, the shading system in office S1 was inactive (manual mode), such that the shading automation system did not change the position of a blind and slat angle regardless of the configuration. Both users were present often at work and spent more than 4 hours on average at work. They changed the position of a blind in their offices when they were present.

As already discussed in [9] and [10], there are several possible reasons why the automation systems fail: (a) sensor failures, (b) control logic problem and (c) longer notification time and delay issues. The data could not identify whether a sensor failure did occur, only capturing the control logic and the delay issues.

On July 3rd, the automation system changed the position of the blind and slat angle when the outside illuminance level measured was higher than 25000 Lux at 13:12 P.M. by the configuration. Likewise, it was changed when the outside illuminance level was lower than 1001 at 19:40 P.M. as it is configured. It seems that the control logic works properly. Regarding the delay issue, it took 2 minutes to deliver Down-trend motion and 6 minutes for Up-trend motion by the automation system. The details of office E1 on 3rd July is illustrated in Fig. 11. Additionally, the light was off and there was not strong wind on the day.

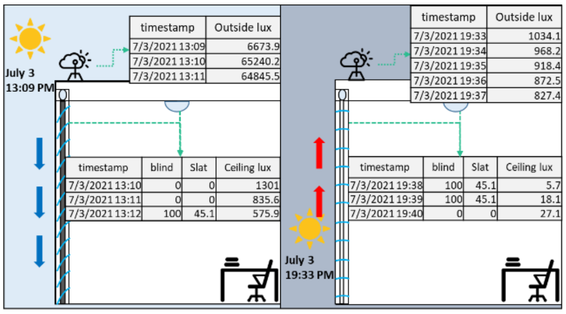

Fig. 11 – Shading automation monitoring in office E1.

The user in office S1 pressed the blind button to open and close the blind on 1st July because the automation system was inactive. A command message generated from a physical button was directly delivered to the blind actuator without delays.

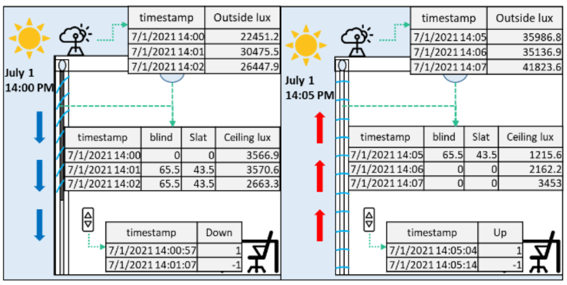

Fig. 12 – Occupant behaviours monitored in office S1.

Both users were in their offices on July 1st, 5th, 6th, 7th and 9th. Comparing the indoor illuminance level measured on the ceiling when both users were in their offices, the user S1 preferred the brighter indoor environment than given by the settings of the automatic system. During 5 days, the average was 759 lux in office E1 and 2409 lux in office S1. Except July 9th as shown in Fig. 13, there is a distinct difference between office E1 and S1. If the automation system in office S1 was active, it would not meet the user's expectations.

Results suggest:

- Control logics of the automation works properly.
- The delay of message delivery exists in automation.
- Centralized automation systems could not meet user's comfort requirements.

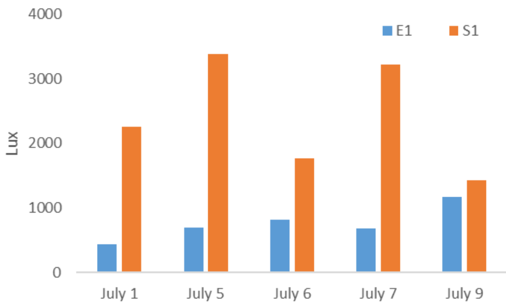

Fig. 13 – Monitored indoor illuminance level on average in office E1 (with automatic mode) and S1 (with manual mode).

3.4 Identifying influential factors affecting user's preference

As mentioned in previous section, the automation system in the south-facing office S1 was inactive and the user changed the position of the blind depending on his comfort preference. An empirical analysis investigates the other associated factors to the user's preference.

Although the shading automation system is configured with outdoor factors (illuminance level and wind speed), it is assumed that indoor climate related factors are the determinants of a user's preference since a user determine whether to open or close a blind according to the indoor climate not outdoor climate. To discover influential factors in office S1, indoor climate related data (e.g. Illuminance level on desk or ceiling) is extracted. By using SPARQL to retrieve related elements and data source information as shown in Frag. 3, sensors and statuses of objects data are easily accessible. Afterwards, the data is filtered based on user's presence at work in order to infer the user's personal indoor climate comfort preference.

Frag. 3 – SPARQL query to get indoor climate related elements and their data source information.

```
SELECT ?system ?elem ?db ?key
WHERE {
```

```
?system brick:hasLocation :S1 .
?system brick:hasPart ?part .
?part brick:hasPoint ?elem .
?elem brick:timeseries ?uid .
?uid brick:storedAt ?db .
?uid brick:hasTimeseriesId ?key .
{?elem rdf:type/rdfs:subClassOf* brick:Sensor}
UNION
{?elem rdf:type/rdfs:subClassOf* brick:Status}
}
```

The indoor climate of office S1 in July was monitored when the user was at work. The total length of the observed period (timestamps) is 9218 based on the measurement in every minute. During his presence, he sent Up-trend command 21 times and Down-trend command 20 times to the blind in his office. These 41 commands include interruptions of the automation system and commands after the interruption. The indoor climate data is labelled as Up-trend when he initiated the Up-trend motion. When he initiated the Down-trend motion, the data are labelled as Down-trend. The remaining 9177 time stamps are labelled as “None”. Before training a model to identify influential indoor factors, pre-processing is required to address data imbalanced. Many techniques [11] to handle this issue have been introduced such as over/under-sampling. Among them, the over-sampling technique is applied to increase the size of the minority class (labelled as Up-trend, Down-trend) to balance the majority class (labelled as “None”). The data with Up-trend and Down-trend labels were duplicated to fit the same numbers of data in each label. After pre-processing, the manipulated length of the timestamps is 18354(9177 labelled as None, 9177 labelled as Up-trend or Down-trend). Based on the processed data, a logistic regression model is applied to predict when the user sends a command.

First, two separate models for Up-trend and Down-trend classifications are trained. To find out whether outdoor climate or indoor climate is more influential to user’s behaviours, the models only with outdoor or indoor climate factors are trained. For evaluation, 70% of the data is randomly selected into a training set and the remaining 30% into a testing set. The illuminance level and wind speed are used for outdoor climate models. The result only with outside factors is in Tab. 3. Although the outdoor climate factors are used in the shading automation system as configuration values, it does not explain the behaviours of the user effectively.

Tab. 3 – Experiment results with outdoor factors

Model		Up-trend	Down-trend
Coefficient	Wind	-0.0052	0
	Illum.	-0.0004	-0.0003
Accuracy	Training	0.57	0.49
	Testing	0.57	0.49

The office S1 has 12 indoor climate factors available. In order to find the important variables, a feature selection technique is applied. In Up-trend motion, illuminance level measured on a desk and the slat angle of a blind are the influential factors among 12 indoor climate factors. In down-trend motion, illuminance level measured on a desk and ceiling and the position of a blind are important. The results of the best models are listed in Tab. 4.

From empirical analysis, the following is inferred for the user in office S1: in Up-trend (blind opening) and Down-trend(blind closing), indoor illuminance level and the status of blind(positon and slat angle) are highly related to his comfort requirements rather than other indoor factors.

Tab. 4 – Results with different indoor factors

Model		Up-trend	Down-trend
Coefficient	Lux(desk)	-0.0172	-0.0001
	Lux(ceiling)	-	-0.0021
	Blind pos.	-	-1.1478
	Slat angle	-0.0018	-
Accuracy	Training	0.85	0.78
	Testing	0.86	0.79

4. Conclusion

In order to integrate multiple and heterogeneous data environments, we built an ontology for building automation data analysis. It explicitly describes the relationships between variables which makes data understandable for both humans and machines. Although each room has a different number of channels monitored and configuration descriptions, the data and relevant information were efficiently retrieved based on the ontology without changing a SPARQL query. In this paper, we focus on convenient access to different data sources for further research with several possible use scenarios. In terms of data engineering, an ontology-based approach provides efficient data handling. This study is limited by the small number of observed data of user’s activities (e.g. blind closing, opening) due to occupants working from home due to COVID-19. Future work will build a scalable model based on the proposed ontology containing a set of evaluated methods and comparisons between them.

5. References

[1] Pritoni M, Paine D, Fierro G, Mosiman C, Poplawski M, Saha A, Bender J, Granderson J. Metadata Schemas and Ontologies for Building Energy Applications: A Critical Review and Use Case Analysis. Energies. 2021 Jan;14(7):2024.

[2] Rasmussen M, Pauwels P, Lefrançois M, Schneider

GF, Hviid C, Karlshøj J. Recent changes in the building topology ontology. InLDAC2017-5th linked data in architecture and construction workshop 2017 Nov 13.

[3] Daniele L, den Hartog F, Roes J. Created in close interaction with the industry: the smart appliances reference (SAREF) ontology. InInternational Workshop Formal Ontologies Meet Industries 2015 Aug 5 (pp. 100-112). Springer, Cham.

[4] Balaji B, Bhattacharya A, Fierro G, Gao J, Gluck J, Hong D, Johansen A, Koh J, Ploennigs J, Agarwal Y, Bergés M. Brick: Metadata schema for portable smart building applications. Applied energy. 2018 Sep 15;226:1273-92.

[5] Noy NF, McGuinness DL. Ontology development 101: A guide to creating your first ontology.

[6] Mahdavi A, Taheri M. An ontology for building monitoring. Journal of Building Performance Simulation. 2017 Nov 2;10(5-6):499-508.

[7] Fierro G. Design of an effective ontology and query processor enabling portable building applications. University of California, Berkeley. 2019 May 10.

[8] Terkaj W, Schneider GF, Pauwels P. Reusing domain ontologies in linked building data: the case of building automation and control. In8th International Workshop on Formal Ontologies meet Industry 2017 (Vol. 2050).

[9] Daissaoui A, Boulmakoul A, Karim L, Lbath A. IoT and big data analytics for smart buildings: a survey. Procedia Computer Science. 2020 Jan 1;170:161-8.

[10] Giraldo-Soto C, Erkoreka A, Mora L, Uriarte I, Del Portillo LA. Monitoring system Analysis for evaluating a building's envelope energy performance through estimation of its heat loss coefficient. Sensors. 2018 Jul;18(7):2360.

[11] Hasib KM, Iqbal M, Shah FM, Mahmud JA, Popel MH, Showrov M, Hossain I, Ahmed S, Rahman O. A survey of methods for managing the classification and solution of data imbalance problem. arXiv preprint arXiv:2012.11870. 2020 Dec 22.

[12] Schneider GF, Pauwels P, Steiger S. Ontology-based modeling of control logic in building automation systems. IEEE Transactions on Industrial Informatics. 2017 Aug 23;13(6):3350-60.

[13] Petrova-Antonova D, Ilieva S. Digital twin modeling of smart cities. InInternational Conference on Human Interaction and Emerging Technologies 2020 Aug 27 (pp. 384-390). Springer, Cham.

[14] Mavrokapnidis D, Katsigarakis K, Pauwels P, Petrova E, Korolija I, Rovas D. A linked-data paradigm for the integration of static and dynamic building data in Digital Twins. InProceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation 2021 Nov 17 (pp. 369-372).