Semantic Building Systems Modeling for Advanced Data Analytics for Energy Efficiency

Filip Petrushevski[1], Stefan Gaida[2], Barbara Beigelböck[1], Miloš Šipetić[1], Gerhard Zucker[1], Christian Schiefer[3], Daniel Schachinger[2], Wolfgang Kastner[2]
[1]AIT Austrian Institute of Technology, Vienna, Austria
[2]TU Wien, Vienna, Austria
[3]Caverion Austria, Vienna, Austria

Abstract

The complexity and variety of today's building systems make analytics of building data a challenging task. To be able to consolidate and analyze the vast amount of information including data regarding system structure, device characteristics, or operational data (e.g. task schedules) as well as monitoring data and external information (e.g. weather data) a reliable semantic model that represents the expert knowledge about building systems is necessary. This paper focuses on the semantic modeling of building systems to support advanced data analytics algorithms for improvement of building energy efficiency.

Introduction

Modern tools for building energy management allow facility managers to access large amounts of data, which represent the energy use of a building. Based on their personal experience, they perform actions to improve the energy performance. However, with the increasing amount of data, it becomes impossible to manually analyze the information and make appropriate decisions. Thus, an automated system that supports decision making based on algorithms for data analytics, simulations and optimization can help facility managers to improve the building performance in a more flexible manner. With advanced data analytics mechanisms, the system can be used not only to visualize monitored data from the past or present and forecast energy consumption, but also to provide feedback on how efficiency measures will improve the performance of the building.

The building process traditionally produces data at different lifecycle stages starting with design and planning, followed by construction and commissioning and finally the operation. The design and planning of building systems is typically done in 2D and not updated consistently for all stakeholders. The resulting well-known problems are collisions during construction phase and incorrectly dimensioned systems. Also, there is a major information gap between the handover from construction phase to operation phase. The information that has been created during planning phase, consisting mainly of structural information of the building and the spatial location of energy systems, is, as indicated, not consistent, and is not made available to the operator of the building. This implies that semantics (e.g. spatial information) that is needed to analyze the operation and efficiency of the building is not available, but needs to be recreated. This information shall be made available in a machine-readable format and augmented with additional information in order to provide a rich source of semantics. An effort to represent the building systems related information as a semantic model used for advanced data analytics for the building operation is presented here.

State of the art

Applying the knowledge of energy efficiency experts through data analytics is becoming an established practice for detecting inefficiency in building performance (Peña et al. (2016)). By using data mining of operational building data, additional knowledge about the building operation could be extracted. Normally, the algorithms are combined with expert knowledge to provide the best insight and make the best decisions. Decision support systems are used for building energy efficiency from the early design phase (Attia et al. (2012)) up to operation (Doukas et al. (2009)). For example, Henze et al. (2015) developed a decision support tool for building energy systems that provides information for building energy use that allows a building operator to quickly distinguish normal and abnormal energy use.

Building modeling methods are used mostly for building design. Existing semantic models of buildings are used to support the collaborative integrated design of buildings (Lee et al. (2016)). Significant work has been done to automate the process of geometric modeling of buildings for the purpose of building thermal energy simulation (van Treeck et al. (2015); Kim et al. (2015)). However, the missing building systems information is hindering their use during operation. To link the design phase information into operation, further methods are needed.

Semantic modeling is generally used to represent the concepts that describe the real world physical systems as precise as possible. At the same time, mapping is provided through utilizing the encoded meaning. This enriches the data to ensure efficient information access and management. The models formally represent domain knowledge which can be very useful for specific data analytics methods such as data mining (Dou et al. (2015)). For example, the building systems expert knowledge can be represented in such a model.

Efforts are being made to link semantic modeling and building information modeling (BIM) (Pauwels and Terkaj (2016)). This comprehensive data model is used during the design phase of the building, but carries the potential to significantly improve building operation as well. Niknam and Karshenas (2015) developed a knowledge based system that automatically accesses information from different sources available as semantic web services (SWS) to provide the information to an energy analysis application. This approach significantly simplifies the fast evaluation of energy efficiency in the early design stage. By combining semantic BIM with rules that encode expert knowledge, a fault propagation approach could be applied to building automation system (BAS) data to implement better fault detection and diagnostics algorithms (Dibowski et al. (2016)). However, since normally not all BAS variables and their semantics are available in a BIM model, a different approach is needed. The necessary developments of the BIM towards building automation are on the way, but not available in the building process yet.

System architecture

In our previous work, we presented an advanced data analytics framework for energy efficiency (Schachinger et al. (2016)). The system architecture describes a system that is capable of fusing the different data sources, which include building operation data, weather data and the additional semantic metadata that describe the building properties. Figure 1 shows its architectural overview. The iterative workflow with its main modules is split into three sections in accordance with the data analytics steps: descriptive analytics, predictive analytics and prescriptive analytics. The interfaces between these steps are defined by the exchanged data artifacts.

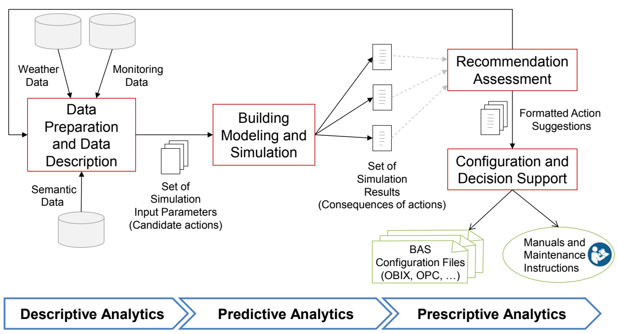

Figure 1: Architecture of the advanced data analytics framework (Schachinger et al. (2016))

An important part of the framework is the data preparation and data description module (Figure 2) which collects data from building systems, together with weather data, as well as additional data about the building (such as building structure, data of the deployed systems, and equipment types) and analyzes them to detect outliers or identify patterns and trends. In this module, based on the descriptive data analytics, candidate actions or improvement measures are generated or selected from a knowledge-base or so-called actions database. Here the actions are described in a format similar to logical expressions. They contain the logic and the parameters that need to be optimized. Possible actions are, for instance, change of set-point temperature, change of schedule, or change of operating modes of HVAC equipment. An example for such an action would be to set the operation schedule of a geothermal heat pump so that more renewable energy is used if its capacity is greater than the building's current demand. The actions are then translated into a set of simulation input parameters which are used by a simulation environment to make an assessment of the effectiveness of each action on the energy performance of the building. This is supported by the semantic description of the building systems to be able to map the correct datapoints. For example, a mapping of the sensor or actuator IDs and addresses could be easily extracted using semantic queries. In this work, we describe the semantic model of building systems used by the advanced data analytics algorithms.

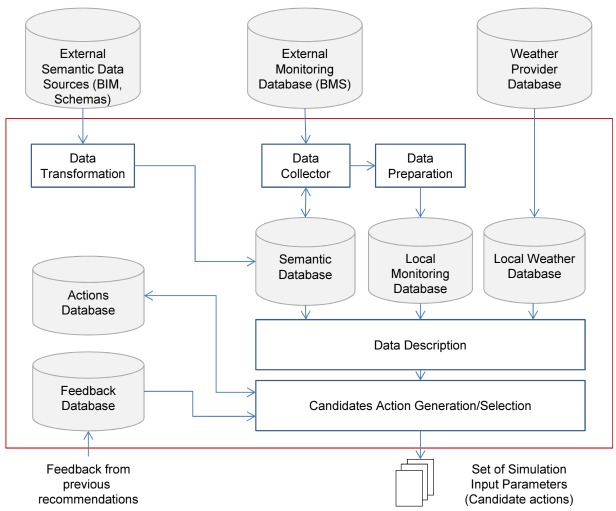

Figure 2: Data preparation and data description module diagram.

Ontology Design

An ontology is a formal and explicit specification of a shared conceptualisation (Studer et al. (1998)). A conceptualisation refers to an abstract model of an aspect of the world by having identified the relevant concepts, their properties and relations. The terms formal and explicit mean that the model can be processed by machines as well as by humans and that types of concepts and the constraints on their use are explicitly defined. Overall, an ontology comprises concepts or classes defining types of entities, instances of these concepts, relations between concepts, and restrictions on relations.

Schachinger and Kastner (2016) present a smart control ontology that enables semantic modeling of building structures, BAS resources and services. As a basis for their work, they reuse already existing ontologies like DogOnt (Bonino and Corno (2008)) or ThinkHome (Kofler et al. (2012)). For this reason, there are also links to concepts of these ontologies which ensures a common understanding. The ontology consists of the four main parts (1) building structure, (2) devices and appliances, (3) data services, and (4) control services. For description of a building structure, different kinds of zones which are relatively arranged to each other are used. Examples for zone types are sites, buildings parts, floors, or rooms. Devices and appliances like sensors and actuators which host data and control services are modeled as various subclasses of the concept *BuildingResource*. A *DataService* represents data that is provided by a service provider (e.g. BAS device). Instances of the concept *ParameterConfiguration* are used for semantic description of the data. Such a configuration has exactly two parameter types defining the meaning of the data values. Examples for parameter types are *Temperature*, *Humidity*, or *Time*. Finally, a *ControlService* is intended to influence these parameters. In favor, at least one *ParameterVariation* describes how and under which conditions a parameter can be changed. Besides the ontology, a well-defined control interface that manages ontology access from connected systems is defined, e.g. a BAS or a building management system (BMS). The interface is mainly based on the Resource Description Framework (RDF) and the WebSocket protocol.

Even though the presented ontology already provides a wide range of concepts for semantic modeling of building systems, some extensions are necessary to support advanced data analytics methods for improvement of building energy efficiency. All new concepts, properties, and constraints are modeled with the open-source ontology editor Protégé[1]. Like the base ontology, also the new one is based on the Web Ontology Language (OWL) 2. In the following, the major enhancements are described in detail. Minor changes like the introduction of sub-devices (or components) are not further discussed in this section.

Rules: The class *Rule* forms the main element to support rule-based fault-detection. Such a rule consists of one condition and at least one conclusion. Thereby, the object properties *hasCondition* and *hasConclusion* are used. The actual condition is given as a data property (*formula*) of the *Condition* class. The class *Variable* in combination with the object property *hasVariable* allows references to building resources like devices, constants, data services, and key performance indicators (KPI). Such a variable can be included in formulas (e.g. for referencing the value of a data service) by the use of symbols which are assigned to each instance of the concept *Variable*. The proposed ontology offers classes for two kinds of conclusions, namely *RegularConclusion* and *IncidentConclusion*. The former are defined in the same way as conditions. This means that they consist of an arbitrary formula and optional variables. Incident conclusions also have assigned variables, but instead of a formula so-called anomaly types are used (*hasAnomalyType*). The class *AnomalyType* is intended to describe problems that can occur within the BAS (e.g. some value is rising too fast, or the breakdown of a component is likely). Examples for instances of this class are *FastRaise*, *FastDecline*, *High*, *Conflict*, or *BreakdownLikely*. This allows to define rules which express that a particular problem exists in the system when some condition is met.

KPIs: A KPI like the coefficient of performance of a heat pump or an air conditioning unit can be represented in the ontology by an instance of the class *KeyPerformanceIndicator*. The formula for calculation of the current value of a KPI is modeled by a data property and variables as previously described for rule conditions. In addition, it is possible to specify an expected value which can be a range or an exact value (*hasExpectedValue*). The object property *concerns* allows to assign the KPI to some building resource.

Incidents: Detected problems during rule-based fault-detection, KPI calculation, as well as by other methods (e.g. statistics-based fault-detection) are represented in the ontology as *Incidents*. Such an incident has at least a data property *time* (date and time with timezone), an anomaly type, and one or more variables. The variables are used to specify the service, building resource, or KPI which caused the incident. In addition, an absolute or relative value and a reference to a rule can be assigned. The value is intended for use with anomaly types where the extent of the problem is an important information (e.g. *High* or *HighDeviation*). The detected incidents will be used to improve energy efficiency of a building in future work.

Model implementation

Figure 3 shows a modeling example of a simple rule with an *IncidentConclusion*. White boxes represent classes and the ones with gray background individuals. In the example, the namespace *ex* is used for all properties. The modeled rule states that active cooling and heating at the same time results in a *Conflict*. For the definition of the formula, JavaScript syntax is used. For the sake of clarity, some classes, instances,

[1] http://protege.stanford.edu/

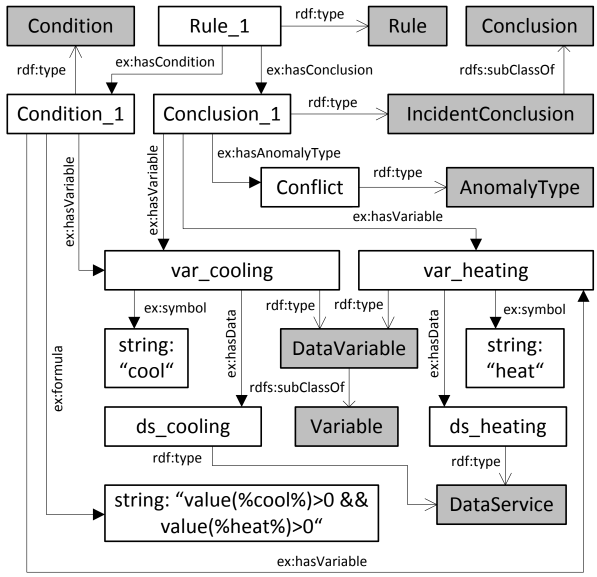

Figure 3: Rule modeling example

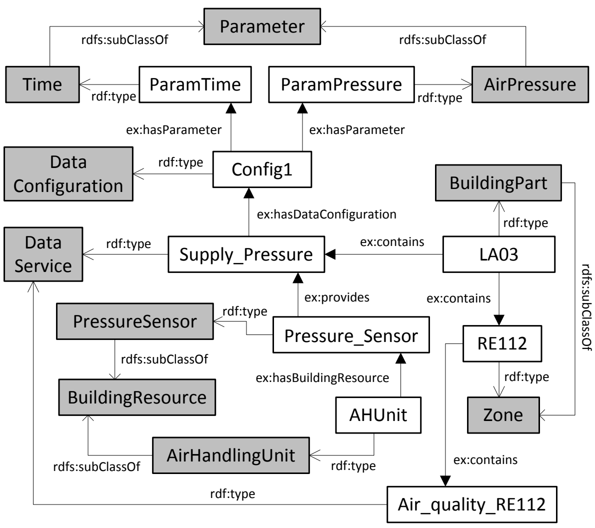

Figure 4: Basic building systems ontology.

and properties are not shown in the figure. This applies especially to the referenced data services.

The semantic building systems model represented as ontology with its classes, relations and properties is shown in Figure 4. The ontology can represent most of the typical building systems with their structure, device characteristics or subsystem dependencies, and is not limited to only one building. Here a part of the implementation of a ventilation system of a commercial building is shown as an example. Figure 5 gives an overview of the topology of the same ventilation system with one of its air handling units (AHUs) and the relevant sensors and actuators represented as building resources and their provided services, as well as separate rooms and their groups represented as zones. As an input for the instantiation, mainly information from the BMS design specification and lists of data-points was used.

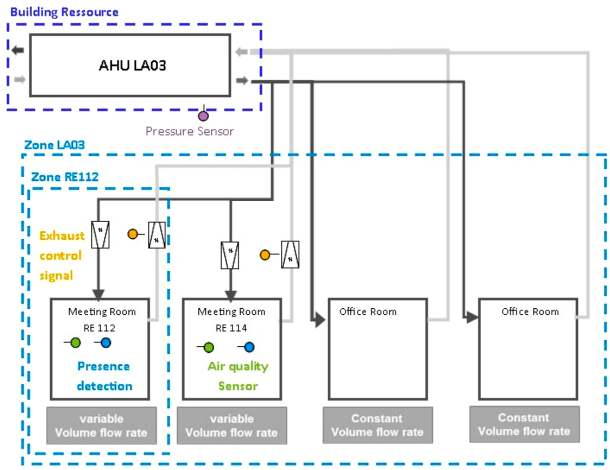

Figure 5: Topology of a ventilation system.

The following SPARQL query[2] is an example of how the semantic model can be used in practice to easily query the variables and their associated data services that are used in the conditions of rules for rule-based fault detection:

```
SELECT ?var ?symbol ?rule ?service ?config
    ?timeparam ?valueparam WHERE
{
    ?rule rdf:type ex:Rule.
    ?rule ex:hasCondition ?cond.
    ?cond ex:hasVariable ?var.
    ?var rdf:type ex:DataVariable.
    ?var ex:symbol ?symbol.
    ?var ex:hasData ?service.
    ?service ex:hasParameterConfiguration
        ?config.
    ?config ex:hasParameter ?timeparam.
    ?config ex:hasParameter ?valueparam.
    ?timeparam rdf:type ex:TimeParameter.
    FILTER NOT EXISTS { ?valueparam rdf:type
        ex:TimeParameter .}
}
```

The result of such a query is used by the data analytics algorithms. Furthermore, similar queries can be used for automatic simulation parameterization in case of predictive analytics.

Results

When it comes to descriptive analytics, for example, in the case of rule-based fault-detection, several rules could be defined in the ontology to automatically detect building performance inefficiencies. The following rules are given as further use cases to show the feasibility of the approach.

Rule 1: Detection of comfort violation

Rule 1 is able to detect comfort violation when it comes to air quality in a particular room and to raise an alarm if the CO_2 levels are above the maximum (e.g. 1000 ppm) defining the comfort band for the air quality. In the example, the rule is coupled with the data service "Air quality" for room "RE112" (averaged over the previous 60 minutes for each timeseries sample):

[2] https://www.w3.org/TR/sparql11-query/

```
condition:
"avg(previous_t($Air_quality_RE112$,60)) >
    1000"

conclusion:
"raise_alarm(High,$Air_quality_RE112$)"
```

When the rule is applied to the monitored data, , several faults are detected. Figure 6 shows these faults for one selected week. On Monday, Tuesday, Wednesday and Friday during different time of the day the hourly average CO_2 concentration was above 1000 ppm and during these periods an alarm is raised automatically. The rule could be also applied on additional comfort parameters, such as indoor air temperature in order to detect thermal comfort violation, but for simplicity the example given here takes into account only the indoor air CO_2 concentration.

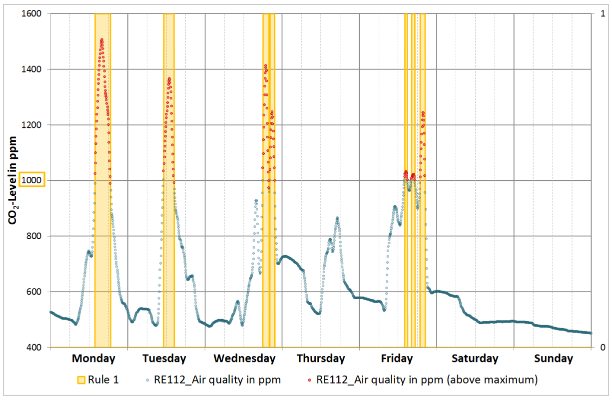

Figure 6: Detection of comfort violation.

Rule 2: Detection of system dimensioning issues

Since Rule 1 is trivial and can only detect comfort violation without pointing the reasons for it, the following rule is extending the conditions and besides "Air quality", data services such as the "Air volume flow" for room "RE112" and the overall "Air supply pressure" are analyzed. With this extension, Rule 2 is able to detect system dimensioning issues such as insufficient volume flow capacity or exceeding maximum occupancy.

```
condition :
"avg(previous_t($Air_quality_RE112$,60)) >
    1000 &&
avg(previous_t($Air_volume_flow_RE112$,60)) >
    80 &&
avg(previous_t($Supply_Pressure$,60)) /
    MAX_Supply_Pressure > 0.8"

conclusion :
"raise_alarm(High,
    "System dimensioning issues detected")"
```

When applied to the monitored data, the issues shown in Figure 7 are detected. The figure shows several sample days when an alarm was triggered. On several occasions on Tuesday, Wednesday and Friday the hourly average CO_2 concentration was above 1000 ppm while the air volume flow for room RE112 and the air supply pressure were on maximum (above

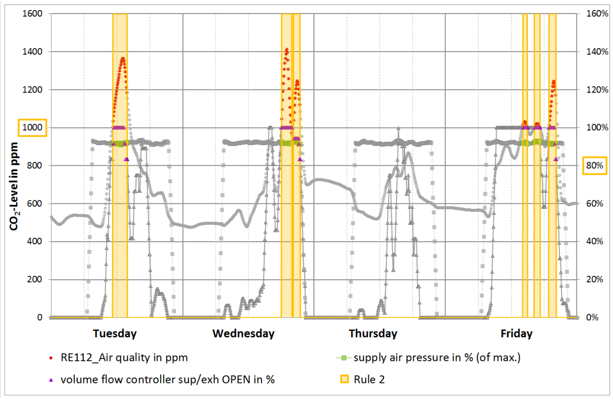

Figure 7: Detection of system dimensioning issues.

80%). This indicates that the maximum occupancy for the room was exceeded (or the volume flow capacity was not sufficient).

Rule 3: Detection of energy inefficiency conflicts

This rule uses data provided by the data service "Presence detection" and "Exhaust control signal" for room "RE112" to raise an alarm if ventilation is running (according to the operating schedule) without detected occupancy.

```
condition:
"max(previous_t($Presence_detection_RE112$,
    60)) == 0 &&
cursor_value($Exhaust_control_signal_RE112$)
    > 5"

conclusion :
"raise_alarm(Conflict,
    $Presence_detection_RE112$,
    $Exhaust_control_signal_RE112$)"
```

If applied to the monitored data, Rule 3 can detect that, for example, on Saturday the ventilation was running (exhaust control signal above 5%) although no occupants were present. Furthermore, on working days the ventilation was operating early in the morning and late in the evening according to a fixed time schedule without occupancy detected. This is shown in Figure 8 for a typical week.

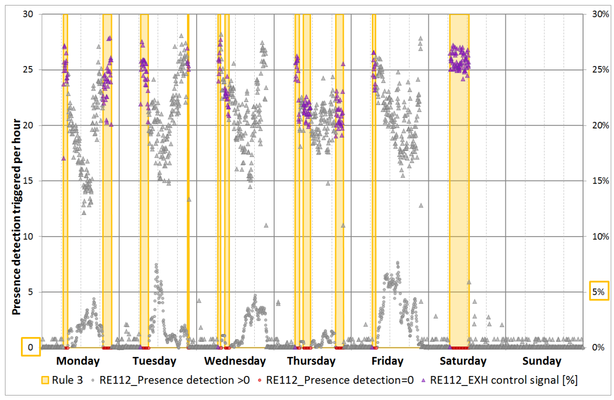

Figure 8: Detection of conflicts that lead to energy inefficiency.

These results show the feasibility of using semantic modeling to link the design phase semantic sources with the operational phase building services.

Conclusion

Today's facility managers face many challenges that not only make it difficult to operate a building efficiently but also to increase the risks for reduced building performance in terms of energy efficiency while keeping the comfort of the occupants. In order to achieve a maximum operational efficiency and cost savings, facility managers must first derive the most comprehensive insights from the collected performance data. This paper shows how a semantic model of building systems and expert domain knowledge represented as rules inside an ontology could provide advanced data analytics algorithms with the necessary information. The results demonstrate the feasibility of the approach to detect faults and performance inefficiencies in building systems such as ventilation. Moreover, the solution could allow energy efficiency experts to institutionalize their knowledge using rule-based systems at the plants they service. For example, an analytics report can guide an on-site maintenance team to choose the best course of action on a daily basis to optimize building operation. Our future work will concentrate on the generation of the candidate actions based on the description of the monitored data such as the presented fault detection and the system knowledge represented in a knowledge-base.

Acknowledgment

This work was funded by FFG (Austrian Research Promotion Agency) under the *ADA-EE* project (849968).

References

Attia, S., E. Gratia, A. De Herde, and J. L. Hensen (2012). Simulation-based decision support tool for early stages of zero-energy building design. *Energy and buildings 49*, 2–15.

Bonino, D. and F. Corno (2008). *DogOnt - Ontology Modeling for Intelligent Domotic Environments*, pp. 790–803. Berlin, Heidelberg: Springer Berlin Heidelberg.

Dibowski, H., O. Holub, and J. Rojicek (2016). Knowledge-based fault propagation in building automation systems. In *2nd International Conference on Systems Informatics, Modelling and Simulation (SIMS2016), Riga, Latvia.*

Dou, D., H. Wang, and H. Liu (2015). Semantic data mining: A survey of ontology-based approaches. In *2015 IEEE International Conference on Semantic Computing (ICSC)*, pp. 244–251. IEEE.

Doukas, H., C. Nychtis, and J. Psarras (2009). Assessing energy-saving measures in buildings through an intelligent decision support model. *Building and environment 44*(2), 290–298.

Henze, G. P., G. S. Pavlak, A. R. Florita, R. H. Dodier, and A. I. Hirsch (2015). An energy signal tool for decision support in building energy systems. *Applied Energy 138*, 51–70.

Kim, J. B., W. Jeong, M. J. Clayton, J. S. Haberl, and W. Yan (2015). Developing a physical BIM library for building thermal energy simulation. *Automation in construction 50*, 16–28.

Kofler, M. J., C. Reinisch, and W. Kastner (2012). A semantic representation of energy-related information in future smart homes. *Energy and Buildings 47*, 169 – 179.

Lee, Y.-C., C. M. Eastman, and W. Solihin (2016). An ontology-based approach for developing data exchange requirements and model views of building information modeling. *Advanced Engineering Informatics 30*(3), 354–367.

Niknam, M. and S. Karshenas (2015). Sustainable design of buildings using semantic bim and semantic web services. *Procedia Engineering 118*, 909 – 917.

Pauwels, P. and W. Terkaj (2016). {EXPRESS} to {OWL} for construction industry: Towards a recommendable and usable ifcowl ontology. *Automation in Construction 63*, 100 – 133.

Peña, M., F. Biscarri, J. I. Guerrero, I. Monedero, and C. León (2016). Rule-based system to detect energy efficiency anomalies in smart buildings, a data mining approach. *Expert Systems with Applications 56*, 242–255.

Schachinger, D., S. Gaida, W. Kastner, F. Petrushevski, C. Reinthaler, M. Sipetic, and G. Zucker (2016). An advanced data analytics framework for energy efficiency in buildings. In *Emerging Technologies and Factory Automation (ETFA), 2016 IEEE 21st International Conference on*, pp. 1–4. IEEE.

Schachinger, D. and W. Kastner (2016, June). Semantics for smart control of building automation. In *2016 IEEE 25th International Symposium on Industrial Electronics (ISIE)*, pp. 1073–1078.

Studer, R., V. R. Benjamins, and D. Fensel (1998, March). Knowledge engineering: Principles and methods. *Data Knowl. Eng. 25*(1-2), 161–197.

van Treeck, C., R. Wimmer, and T. Maile (2015). Bim für die energiebedarfsermittlung und gebäudesimulation. In *Building Information Modeling*, pp. 293–303. Springer.