DEVELOPMENT OF A CONCEPTUAL DATA MODEL FOR THE DIGITAL DESCRIPTION OF FAULTS IN THE BUILDING SYSTEMS´ OPERATION

P. Alfonso[1], T. Rist[1], T. Müller-Eping[1], N. Réhault[1] and P. von Both[2]
[1]Fraunhofer Institute for Solar Energy Systems ISE, Freiburg, Germany
[2] Karlsruhe Institute of Technology, Karlsruhe, Germany

ABSTRACT

Malfunctioning of Heating, Ventilation and Air Conditioning (HVAC) systems causes user discomfort and increases energy consumption. Fault detection and diagnostics (FDD) methods can contribute reducing energy losses by early detecting faulty operations of HVAC systems and thus supporting corrective actions. However, Model-based FDD methods require a vast amount of data for set-up, such as time-series data from sensors and actuators and semantic data documenting technical properties of the supervised systems. This paper presents the conceptual data model for a database, in which we aim to collect, describe, and classify semantic information about operational faults of HVAC systems, the buildings, the HVAC systems themselves including their data points (sensors/actuators). The main objective is to provide a semantically unified structure for the documentation of faults in HVAC systems to enable developers and engineers training, testing and comparing FDD methods.

INTRODUCTION

Several studies have shown that buildings consume more energy than originally planned. The difference between the energy performance target values and the actual energy consumption is known as “performance gap” (Kallab et al. 2017; Rory Jones and Pieter de Wilde 2015). Malfunctions in the operation of the Heating, Ventilation and Air Conditioning Systems (HVAC) are one of the factors that contribute to this phenomenon. System faults that remain undetected over large periods of time increase the energy consumption and cause financial losses. In the last decades, researchers have shown that significant energy savings could be achieved by early detecting the faulty operation of HVAC systems by using Fault Detection and Diagnostics (FDD) methods (Kim and Katipamula 2018).

FDD methods can be based on different mathematical models, each of which has its advantages and drawbacks (Kim and Katipamula 2018). The literature highlights the absence of a systematic testing and benchmark for FDD methods, which hinders the selection of the most appropriate FDD approach for each concrete real-life situation (Benndorf et al. 2018; Li and O’Neill 2018). In most of the cases, researchers use simulated faults to train and test the FDD methods, and evaluate the impact of the faults in the systems (Cui and Wang 2005; Yan Chen et al. 2018; Hong Tianzhen 2016; Yanfei Li and Zheng O´Neil 2016). However, Bruton et al. express the importance of testing the FDD tools with real data sets to increase their versatility (Bruton et al. 2014).

The set-up of FDD methods requires time series data from sensors and actuators, e.g. from Building Automation Systems (BAS) protocols and semantic information. Semantic data includes data point names, as defined in the BAS protocol and description (e.g.: name: “PG-344 Centrale Frigo\Temp. Uscita Torre TE01C” and description: “Tower Out TemperatureTE01C”); building architectural and structural aspects (e.g.: “ADR XX T1”); HVAC systems (e.g.: “TE01C” for the Cooling Tower); fault descriptions (e.g.: *Fault001*: Condenser Water Return Temperature Sensor Failure: Value Too High); and the FDD method (e.g.: type: rule-based, applied formula, etc.). Unfortunately, these data usually has no standardized schema or is incomplete. Without a unique defined structure, analysts expend a great effort in manually acquiring and understanding the information appropriately.

Researchers have done many attempts to structure the data in buildings by means of interoperable share knowledge resources, like standardized workflows and metadata standards. One of the most important metadata schemas is Industry Foundation Classes (IFC). It is an industry-specific data model, which is largely used in the building sector for the digital description of buildings in Building Information Modeling (BIM) (BuildingSmart 2020). An innovative approach is Brick schema (Brick Community 2020), which gathers contextual data of

DOI: 10.3217/978-3-85125-786-1-62

buildings, systems, data points and their relations providing a concrete ontology (Balaji et al. 2016). It was built upon the idea of “tags” from Project Haystack (Project Haystack Corporation 2020). However, these schemas focus on the description of buildings, systems or/and data points and do not propose a model to describe HVAC operational faults. The challenge to build a fault database is to design a model that documents both: the semantic information of the building, its systems and data points, together with the FDD methods and the operational faults occurring in the HVAC systems.

This paper presents a conceptual data model for the description of faults in HVAC operation. For the conceptual model, we did not define all the specific terms and concepts to use in the structure of the description. We identified the domains of information that we need to describe the fault and how these groups relate to each other. In a second step, we developed an ontology and taxonomy for the model to describe the faults and FDD methods. With the ontology and taxonomy we identify, distinguish and give hierarchy to the specific terms and concepts to be used in each domain.

The information is encoded as semantic triples in which each semantic statement has the format: subject–predicate–object (e.g.: “Fault001 isLocatedIn TE01C”). The conceptual data model is the ground foundation of a Triplestore or Resource Description Framework store (RDF store) for the collection of semantic information of HVAC operational faults. The objective is to use this structured information to parameterize and systematically test FDD methods.

In this article, we first describe the conceptual data model, highlighting the information groups and relations. Then we explain the ontology and taxonomy (normalized vocabulary and its hierarchy) developed to describe the operational faults. Afterwards, we give an example of a fault description using the ontology. Finally, we conclude and propose additional work.

STATE OF THE ART

Collecting and interpreting buildings semantic data involve labor intensive effort by experts. Finding a complete schema to describe the vast variety of buildings, systems and data points remains still a challenge. A data schema should at minimum be able to model the buildings architectural and structural aspects, the systems technical data and the relevant data points. In the following, we provide an overview on the state-of-the-art of the last years on semantic schemas developed for digital building descriptions.

Project Haystack (Project Haystack Corporation 2020) is an open source initiative to standardize semantic data models. By providing a common taxonomy, Haystack defines “tags” that are associated to the entities, such as an air handling unit or a data point. The vocabulary (metadata terminology in the “tags”) describes the building equipment, data points and properties without changing the entity name. Entities are linked through references “tags”.

IFC allows a digital description of the building during its whole lifecycle. A wide range of software support *.ifc* files. It captures the buildings structure and the relationship of the building components. The later version allows modeling sensors, actuators and controllers. The data format/standard “Construction Operations Building Information Exchange” (COBie) supports IFC with simpler export formats (E. William East 2007). The Green Building XML schema is another BIM exchange format with focus on the energy performance analysis (gbXML.org 2020).

The World Wide Web Consortium (W3C) proposes a formal language and rules to handle data on the Web. Their idea of a vocabulary is a group of defined concepts, domain hierarchy and definitions to help the data interaction (W3C 2020). For example, the Semantic Sensor Network ontology (SSN) (OGC & W3C 2020) describes sensors with focus on their measurements, their observation and methods. However, it lacks of formal language to describe the location, units or features, but it can be linked from other ontologies. The project Smart Appliances REFerence (SAREF) unified different ontologies to be used in smart appliances (Frank den Hartog et al. 2015).

Bhattacharya et al. (2015) analyzed in their paper the three metadata schemas for buildings: Project Haystack, IFC and the Semantic Sensor Web through the SAREF project. The paper evaluates the capacity of the schemas to model the contextual, spatial and functional relationship between the sensors and the building environment (Bhattacharya et al. 2015). The results show that none of the three schemas captures all the required data. Based on this finding, Brick developed a schema reusing some concepts of the above mentioned schemas, and defining a new ontology to represent the sensors, systems and the relationship among them (Balaji et al. 2016).

We developed our model following the work done in Brick and focusing on the documentation of the faults in the building systems operation and the FDD methods. The completeness of the Brick´s taxonomy and the flexibility of the Brick´s ontology allow us to combine Brick with our ontology, which describes the faults and the FDD methods. We used the Brick´s ontology to model the HVAC systems and the data point; and developed our own ontology for description of the faults and FDD methods.

CONCEPTUAL DATA MODEL OF OPERATIONAL FAULTS

The conceptual data model for the digital description of the operational faults focuses on capturing the semantic data and relationships of buildings architectural and structural aspects, HVAC systems, data points, FDD methods and faults. Real data and results from previous projects are the ground information to drive the basic requirements of the conceptual data model. As an example for this paper, a typical fault scenario denoted by “Fault001: Condenser Water Return Temperature Sensor Failure: Value Too High” is introduced. In the next section, we show how our ontology is used to describe this fault.

The diversity of building systems and their available metadata require a flexible model that deals with uncertainty and expansion. A triple structure states that a subject (resource) has a relationship predicate (relationship) with an object (resource or literal) (Ora Lassila and Ralph R. Swick 2020). This structure has been proved to correctly represent large, interconnected amounts of data. By using this concept we also benefit from already developed standards: RDF, RDF schema (RDFS), Web Ontology Language (OWL) and Brick. Reusing them allows us to focus on developing unified and structured data model to digitally describe HVAC systems operational faults. From our current knowledge, such data model is still inexistent.

We defined five domains of information related to the detected operational faults: Building, HVAC Equipment, Data Point, FDD method and Fault (Figure 1). Within the five groups, it is possible to collect all the relevant information to document the fault. Afterwards, we aim to query the semantic information form the database to label the time-series data as faulty or fault-free.

The labeled data, together with the FDD parameters (e.g.: temperature thresholds of a rule-based system), data point structured description (e.g.: supply air temperature sensor and supply air temperature set point), HVAC technical information (e.g.: nominal power consumption), and building structure (e.g.: supplied zones) parameterize the FDD method, which can be an already existing method or a new one. In case of evaluating a new FDD method, its results - the detected faults in a system - can be compared against the fault documented in the database for the same system, to measure its accuracy.

- **Building:** captures the descriptive semantic information of the architectural and structural aspects of the buildings, e.g.: allocation of the space, rooms and zones. Other properties like year of construction, ground floor and geolocation are included.
- **HVAC Equipment:** captures descriptive semantic information about the HVAC equipment and their technical data. The relationships with the other groups model its location, its related data points, and its detected faults. Also, it relates to itself to model its subsystems.
- **Data Points:** captures the name of the data points and characteristics, like unit of de measure values (e.g.: kW, °C).

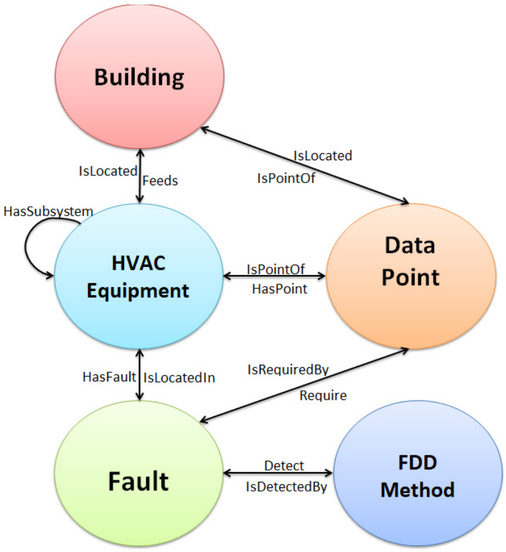

Figure 1: Information domains and their relationship

- **FDD Method:** captures descriptive semantic information about the FDD method (e.g. type: black box, grey box), and the details used to detect each fault (parameters, ranges, formulas, etc.).
- **Fault:** captures descriptive semantic information about the detected operational faults, together with the time range of the occurrence. The next section details its taxonomy and ontology.

Each domain is designed using an arch-node graph, and the statements follow a triple structure. The model has resources and attributes. The resources are the subjects and objects of the triple. The attributes include the data properties as objects of the triple and relationships as predicates of the triple. The objects of the triple act as “*tags*”. A “*tag*” contains a semantic term or concept (a word). A “*tag*” or group of “*tags*”, linked to a subject, provides the semantic description for the subject. This concept follows the idea of “*tags*” and “*tag-sets*” from Project Haystack and Brick (Project Haystack Corporation; Brick Community). Providing a defined vocabulary for the “*tags*” (ontology and taxonomy), it is possible to document the detected operational faults in a standardized way.

ONTOLOGY AND TAXONOMY

An item "to tag" (subject in the triple) relates (predicate in the triple) to a "tag" (object in the triple). The items "to tag" are elements in the database that we want to describe (e.g.: Fault001). All the "tags", linked to the same resource, structure the description of item "to tag". Each "tag" contains a concept. A concept is a piece of semantic data used to describe the items "to tag" (e.g.: "Fault", "Supply", "Cooled_Air", "Sensor_Failure", "Pressure"). We defined a minimum of concepts needed to describe the operational faults. We use a group of four "tags" to describe the faults ("*four tag system*").

The taxonomy and ontology provide the formal language for the "tags" and the hierarchy structure for the vocabulary. Individual terms describing the same characteristic are grouped into a global concept. Global concepts and items "to tag" are the classes in the ontology. The individual concepts are the subclasses.

Four tag system

Supplementary to the operational faults found in different HVAC systems, we analyzed operational faults from the literature (Roth et al. 2004; Cheung and Braun 2015). From them and based on the "tag" concept we built a "four tag system" (Figure 2). The four classes for the "four tag system" are:

- "***Position tag***": Indicates the position of the fault (e.g. supply, room). Additional positions like "in" or "before" are available. In these cases it is necessary to specify the equipment related to the position.
- "***Medium tag***": Indicates the physical medium transporting the energy (e.g.: Water, Air).
- "***Measurement tag***": Indicates the variable being measured (e.g.: temperature, volume flow). In case the fault affects the relation between two or more systems (e.g.: asynchronous operation), all the affected measures are considered (e.g.: "*Operation_Status>Fan*" and "*Operation_Status>Condenser> Pumps*")
- "***Fault description tag***": Defines four distinctive faults´ types. Additional data properties allow adding details to the description.
    - *DP fault*: there is a fault in the measure of the sensors or commands.
    - *Target value/limit:* the target value or set point is exceeded or not reached.
    - *Alarm:* an alarm indicates a malfunction in a system.
    - *Mechanical or operation element fault:* there is a mechanical, electrical or electronic problem in some of the HVAC component, without including the DP (e.g.: "Supply Air Fan Return Air Fan Mismatch" or "Cooling Coil Hydronic Valve Hunting").

Figure 2 shows a schema of the "four tag system". First, an item "to tag" (e.g: Fault001) is model as rdf:type with the "tag": "Fault". Then, each "tag" ("Return", "Condernser_Air", "Sensor_Failure", "Temperature") from the "four tag system" relates to the item, adding the semantic information needed to describe the fault. Finally, the "tags" relates further to other concepts, extending the description.

Taxonomy and hierarchy

Basic terms commonly use at our labs provide the vocabulary for the ontology (Figure 3). For the fault ontology, the "Position", "Medium", "Measurement" and "Fault description" define the classes in the ontology. The classes and subclasses have data properties, like value and unit, to add details. The data properties and relations are set at a resource level (subject and object of the triple), which means that the "tag" (concept) relates directly with the item "to tag". Additionally, as each tag is independent from each other, it increases the flexibility in the description, allowing the model to deal with uncertainties and novelties. For example, in the case the position of a sensor is unknown (uncertainty), the "Position tag" can be avoided or added afterwards. Moreover, each item "to tag" can have several "tags" from the same class. Also, we considered an "undefined_tag" in each class, for concepts that are not considered in the ontology.

Additionally to the fault ontology, we developed an ontology to describe the FDD methods and their parameters. The applied taxonomy uses concepts extracted from the bibliography (Srinivas Katipamula and Michael R. Brambley 2005a, 2005b).

We differentiated between automatic and manual methods. In the manual method, the expert detects the faults by manually analyzing the information (expert knowledge). For the automatic methods, model-based and rule-based algorithms are used to detect the faults with more or less human intervention. Here, the model-based FDD algorithms are subdivided into two different approaches which are based on the comparison of the current plant measurements either with the data given by a simulation model describing the nominal behavior of the system or with historical plant measurements that have been identified as normal operation by expert knowledge. Additionally a "*hybrid_tag*" is added for combined methods.

Depending on the FDD method, the required parameters vary, in which case only the "Limit" and "Threshold" are considered as subclasses of "FDDParameters". The "FDDParameters" class includes the data properties: name, value, unit and description, which allows modeling a wide range of information.

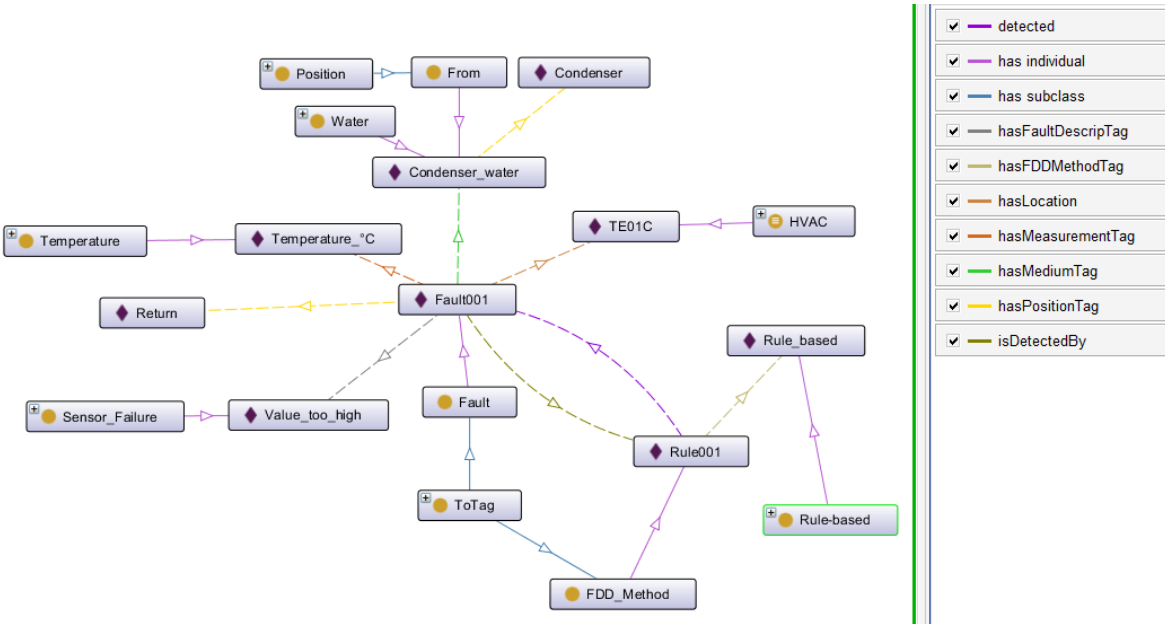

Figure 2: “Four tag system” schema

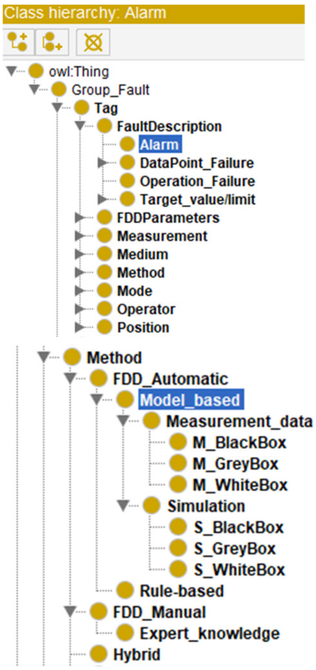

Figure 3: Fault and FDD method ontologies in Protégé (Stanford University 2020)

RESULTS

The database aim to provide researchers with real and structured descriptions of the building systems operation faults. The structured semantic data aim to reduce the time consumed nowadays in processing the information to set up the FDD methods.

Using the resources as “tags” allows the user to maintain the faults, data points and systems´ names, linking it with predefined concepts and additional unstructured semantic data. It increases the flexibility of the schema, but at the same time allows developing automatic queries to retrieve the data.

Through the “Fault description tag”, it is possible to identify the cause/type of fault. These data is used to train a model-based FDD method, adding the fault type to the fault occurrence and the component malfunctioning.

Our study also shows that Brick performs better than other data schemas, when modeling data points. We selected Brick, together with our ontology, to populate the data base.

FUTURE WORK

Next steps include linking Brick with our ontology to model the HVAC systems and its data points.

Before developing the database, future works will focus on testing the ontology by using data from several commercial buildings. The results will show if the taxonomy models the HVAC equipment and document their faults in their totality.

After ensuring the ontology, we will start working on linking it to the FDD methods for their parameterization and comparison.

SUMMARY

Researches show that although the use of FDD methods has increased in the last decades, there is still a gap for a systematic testing and benchmark for FDD methods. One of the reasons is the lack of a common framework and available data, like fault identification, to train and compare the methods. Experts consume high amount of time and effort to collect and process the semantic information, to set-up the methods,

DOI: 10.3217/978-3-85125-786-1-62

which makes the FDD training and comparison not economically profitable in some cases.

Experts have done many attempts to develop semantic data schemas to increase the interoperability of data exchange (Brick Community 2020; Project Haystack Corporation 2020). However, none of the current developed schemas provide a taxonomy for the description and the classification of faults in HVAC systems.

As an attempt to fill this gap, here we presented a conceptual fault data model to digitally describe and classify faults detected in the operation of HVAC systems. The conceptual fault data model builds the basis for the development of a fault database (RDF store).

Additional to the model, we developed an ontology and a taxonomy. We chose a "tag" approach, similar to Project Haystack and Brick. The objects in the triple statement act as "tags". Each "tag" provides a semantic concept; the linked resources define the description of an item (e.g.: fault) without changing its name.

We aim to provide a source of semantically unified data of documented faults in HVAC systems. The data will enable developers to set up and compare FDD methods.

Future work includes linking the developed fault ontology with Brick to complete the ontology and taxonomy for the development of the data base.

PUBLICATION BIBLIOGRAPHY

Balaji, Bharathan; Agarwal, Yuvraj; Berges, Mario; Culler, David; Gupta, Rajesh; Kjærgaard, Mikkel Baun et al. (Eds.) (2016): Brick: Towards a Unified Metadata Schema For Buildings. BuildSys '16,. Palo Alto, USA.

Benndorf, Gesa A.; Wystrcil, Dominik; Réhault, Nicolas (2018): Energy performance optimization in buildings: A review on semantic interoperability, fault detection, and predictive control. In Applied Physics Reviews 5 (4). DOI: 10.1063/1.5053110.

Bhattacharya, Arka; Ploennigs, Joern; Culler, David (Eds.) (2015): Short Paper: Analyzing Metadata Schemas for Buildings: The Good, the Bad, and the Ugly. BuildSys'15. Seoul, South Korea.

Brick Community (2020): Brick schema. Available online at https://brickschema.org/.

Bruton, Ken; Raftery, Paul; Kennedy, Barry; Keane, Marcus M.; O'Sullivan, D. T. J. (2014): Review of automated fault detection and diagnostic tools in air handling units. In Energy Efficiency 7 (2). DOI: 10.1007/s12053-013-9238-2.

BuildingSmart (2020): Building Information Modeling (BIM). Available online at https://www.buildingsmart.org/.

Cheung, Howard; Braun, James E. (2015): Development of Fault Models for Hybrid Fault Detection and Diagnostics Algorithm. NREL.

Cui, Jingtan; Wang, Shengwei (2005): A model-based online fault detection and diagnosis strategy for centrifugal chiller systems. In International Journal of Thermal Sciences 44 (10). DOI: 10.1016/j.ijthermalsci.2005.03.004.

E. William East (2007): Construction Operations Building Information Exchange (COBIE): Requirements Definition and Pilot Implementation Standard. US Army Corps of Engineers, Engineer Research and Development Center.

Frank den Hartog; Laura Daniele; Jasper Roes (2015): Study on Semantic Assets for Smart Appliances Interoperability. SAREF. TNO.

gbXML.org (2020): Green Building XML. Available online at https://www.gbxml.org/.

Hong Tianzhen (Ed.) (2016): Modeling and simulation of operational faults of HVAC systems using ENERGYPLUS. SimBuild´16. Salt Lake City, UT.

Kallab, Lara; Chbeir, Richard; Bourreau, Pierre; Brassier, Pascale; Mrissa, Michael (2017): HIT2GAP: Towards a better building energy management. In Energy Procedia 122. DOI: 10.1016/j.egypro.2017.07.399.

Kim, Woohyun; Katipamula, Srinivas (2018): A review of fault detection and diagnostics methods for building systems. In Science and Technology for the Built Environment 24 (1). DOI: 10.1080/23744731.2017.1318008.

Li, Yanfei; O'Neill, Zheng (2018): A critical review of fault modeling of HVAC systems in buildings. In Build. Simul. 11 (5). DOI: 10.1007/s12273-018-0458-4.

OGC & W3C (2020): Semantic Sensor Network Ontology. Available online at https://www.w3.org/TR/vocab-ssn/.

Ora Lassila; Ralph R. Swick (2020): Resource Description Framework (RDF) Model and Syntax Specification. Available online at https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/.

Project Haystack Corporation (2020): Project Haystack. https://project-haystack.org/.

Rory Jones; Pieter de Wilde (Eds.) (2015): The gap between simulated and measured energy performance: A case study across six identical new-build flats in the UK.

Roth, Kurt W.; Westphalen, Detlef; Llana, Patricia; and Feng, Michael (Eds.) (2004): The Energy Impact of Faults in U.S. Commercial Buildings. International Refrigeration and Air Conditioning´04. West Lafayette, In.

Srinivas Katipamula; Michael R. Brambley (2005a): Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems— A Review, Part I. In HVAC&R RESEARCH 11 (1).

Srinivas Katipamula; Michael R. Brambley (2005b): Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems— A Review, Part II. In HVAC&R RESEARCH 11 (2).

Stanford University (2020): A free, open-source ontology editor and framework for building intelligent systems. Protégé. Available online at https://protege.stanford.edu/.

W3C (2020): SEMANTIC WEB. Vocabulary. Available online at https://www.w3.org/standards/semanticweb/ontology.

Yan Chen; Sen Huang; Draguna Vrabie (Eds.) (2018): A simulation based approach for impact assestment of physical faults: large commercial buildings HVAC case study. Building Performance Modeling´18. Chicago.

Yanfei Li; Zheng O´Neil (Eds.) (2016): An EnergyPlus/OpenStudio based Fault Simulator for Buildings. ASHRAE Winter Meeting. Orlando, Fl.