# Automated classification of HVAC systems through analysis of system behaviour

*Björn* Kämper [a], *Maximilian* Both [a], *Nicolai* Maisch [a], *Jochen* Müller [a]

[a] Institute of Building Services Engineering, TH Köln, University of Applied Sciences, Cologne, Germany, { bjoern_klaus.kaemper; maximilian_alexander.both; nicolai.maisch; jochen.mueller}@th-koeln.de.

**Abstract.** Technical monitoring applications support the operator in identifying potential improvements in plant operation and deriving recommendations for action. Today, the integration of building automation network data points into technical monitoring applications is complex and costly. The current state of the art is to manually integrate data points into dashboard applications. The decision which data point describes which HVAC component is made by engineers. The basis for the decision is usually identifiers or data point descriptions. An automated recognition of the structural information of the HVAC systems would help to automate the process and make the implementation of a technical monitoring in existing buildings and new buildings easier. One way to determine structural information from data points is to look at the behaviour of an HVAC system type.

Common HVAC systems follow known regularities in their construction and behaviour. By analysing the system behaviour, conclusions can be drawn about the heating circuit type (flow temperature controlled, outdoor temperature controlled, etc.). For example, an outdoor temperature-controlled heating circuit will behave differently than a flow temperature controlled heating circuit. Likewise, by analysing the system behaviour, it is possible to predict which data point is assigned to which system component. If two data points increase by a similar value with a short time offset, it is highly probable that they represent the supply and return temperature of a heating circuit. If the flow and return are heated up, it is likely that a pump has started up beforehand, which was represented by a binary switching command.

This paper describes an automated method to classify heating circuit data based on system behaviour. BACnet trend objects of heating circuits from different buildings, that are maintained by the building management of the city of Cologne, serve as data sets. This ensures that the developed method can be applied to a wide variety of heating circuit types. The automated classification of a heating circuit is intended to reduce the effort of manually assigning data points to specially created dashboards. Building operators can thus be supported in the creation and implementation of technical monitoring.

**Keywords.** Building Automation Systems, Building Control Systems, Classification of HVAC systems, Datapoint analysis, BACnet, reducing manual effort
**DOI**: https://doi.org/10.34641/clima.2022.169

## 1. Introduction

In buildings, technical monitoring (TMon) can serve as a central information platform for the operation of building services equipment. With the help of a TMon, the personnel responsible for building services engineering can identify potential for improvement in system operation and derive recommendations for action [1].

Operating information visualized in TMon applications originates from the automation level of the building. In this level the information is mapped to the data points of the communication protocols. Typical in the field of building automation are for example the communication protocols Building Automation and Control Networks (BACnet), Konnex-Bus (KNX) or Local Operating Network (LON). In these common communication protocols, it is not possible to store information about the system structure for individual data points. This means that a data point describing a temperature has no information about the fact that it is describing a temperature and whether the temperature is measured e.g. in a heating circuit or in a water boiler. State of the art is that information about the measured value, the system type or the measurement location is stored within the data point in the form of textual descriptions that can be understood by humans [2].

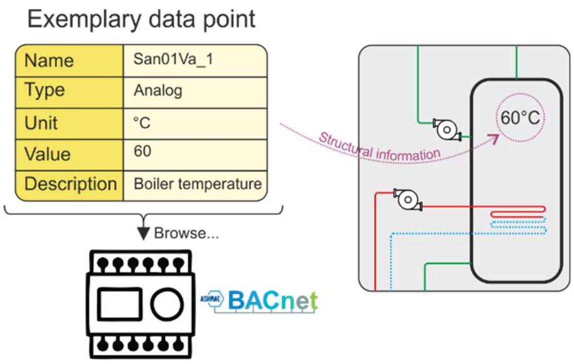


**Fig. 1 -** Exemplary datapoint which stands for the hot water temperature in a boiler. Structural information is only stored as text in the description.

For the creation of a TMon in existing buildings, it is necessary under today's conditions to consider these textual descriptions of the data points individually and to assign the data points manually to the corresponding applications and icons. Since such a manual assignment is associated with high effort and costs, TMon applications are not widespread in building technology, although they are indispensable for the energy- and cost-efficient operation of buildings [3].

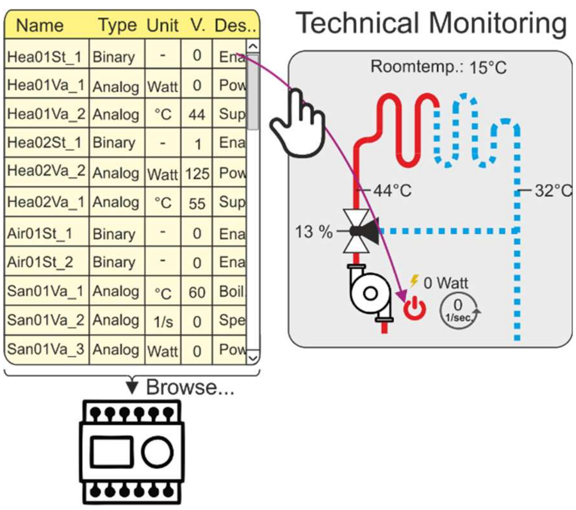


**Fig. 2 -** Manual assignment of a data point to a technical monitoring application.

By linking data points with corresponding structural information, the information content of a data point can be increased. This increased information content can be used to increase the degree of automation in the creation of TMon applications.

Since there are no mandatory attributes for meta-information, structural information, or textual descriptions in current HVAC communication protocols, there may not be enough information to derive structural information from.

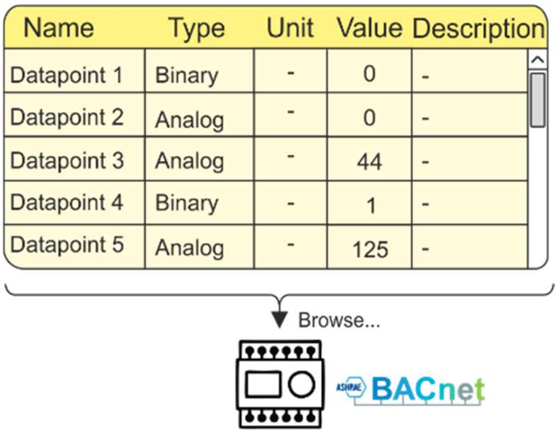


**Fig. 3 -** Example of data points that do not have information about physical units, descriptions, or meaningful labels.

This paper describes how behavioral rules can be derived from real, non-processed, building services engineering data. These behavioral rules are derived from the generally known laws of building services engineering. The goal is to be able to make predictions on the basis of the behavioral rules as to whether an individual data point is a setpoint, a measured temperature or a valve position, for example. In addition, it should be possible to recognize with the help of the behavior rules whether several data points can be assigned to a system component (e.g. a heating circuit).

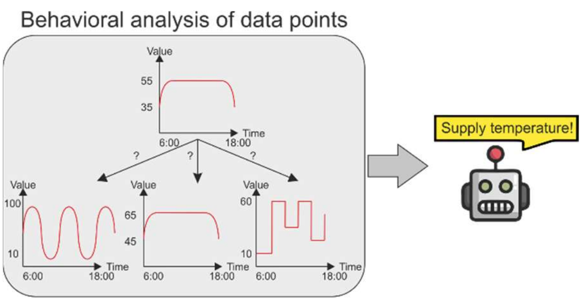


**Fig. 4 -** By the behaviour of a data point, conclusions should be made about structural information.

## 2. State of the art

Today, there are already different approaches how structural information of HVAC data points can be determined or provided:

[4] describes how textual descriptions of HVAC data points can be analyzed using Natural Language Processing (a subfield of artificial intelligence).

[5] extends and improves the previously described approach by considering meta-information such as physical units. Through this analysis, information about the component and/or the installation location can be derived.
Project Haystack is an open-source project that aims to develop a semantic data model for the

operational data in buildings. This semantic data model is primarily developed independently of a technology [6]. In the next steps, these semantic data models are to be mapped to the established communication technologies of building services engineering. The Haystack data model is to be implemented in BACnet, for example, by extending BACnet data points with tags in which additional structural information is stored [7]. Project Haystack thus offers a technological possibility to provide structure information about data points of the HVAC-Systems in the automation level. However, an automated generation and assignment of structure information to data points is not part of Project Haystack. In existing automation environments, these tags would also have to be linked manually.

[8] describes a process how data points from HVAC can be automatically mapped to standards like Project Haystack or Brick Schema. In this process, trend curves and labels of data points are evaluated and classified accordingly with the help of different machine learning algorithms. In the present paper a procedure is described which is similar in principle [8]. However, in this paper only the trend length (value over time) of single data points are considered for the analysis. [8] additionally includes textual descriptions in the analysis.

## 3. Procedure

In order to be able to make predictions about the structural information behind a data point, existing historical data (referred to as training data in the following) of different buildings were analyzed with a Java script. From this training data, commonalities and behavior rules for certain structural components (temperature sensor, valve, pump) were derived. For validation purposes, the commonalities and behavioral rules were tested against a further plant in a final step.

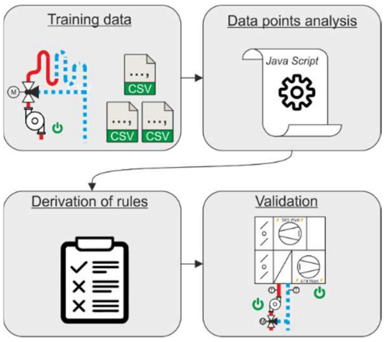


**Fig. 5 –** Procedure for deriving commonalities and behavior rules

### *3.1 Data basis*

The rules were provided from existing and historicized data of buildings of the city of Cologne. The building management of the city of Cologne has connected a large number of buildings to the automation system and histories various building services systems in order to find potential for improvement or to document alarms.

The building management of the city of Cologne uses BACnet for communication and historization of HVAC data points. With BACnet, trend objects can be created for individual data points, which record the corresponding history of a data point [9]. The historization can be implemented with BACnet in two different ways. The values can either be recorded with the polling method and thus sampled at fixed intervals, or with the change-of-value (COV) method, in which a threshold value is defined that determines at which change of the value a recording of the data point should take place. Since COV leads to a smaller amount of data, the city of Cologne historizes the trend data with the COV method for reasons of efficiency.

The historized trend data can be exported by the building management the city of Cologne and made available as a CSV file.

Historic data from ten heating circuits from different buildings were used to derive the behavior rules. The heating circuits have different data points (some have recorded the supply temperature and others the return temperature) and different control strategies (supply temperature controlled, weather temperature controlled). The historized data points are from both summer and winter months. In total, over 700,000 recorded value pairs were considered.

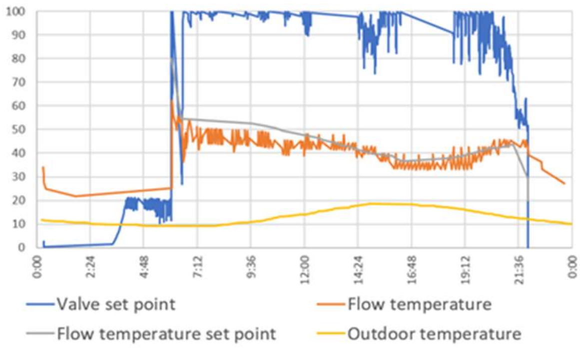


**Fig. 6 -** Exemplary trend curves of a heating circuit over a period of 24 hours.

Before the script examines the historized data points for commonalities and dependencies, the available data still had to be edited. During the historization of the data points, status messages are sometimes recorded instead of the value due to the system. These individual logs had to be removed from the historized data. This could be done with a simple function that deletes all logs that do not contain pure numerical values.

### *2.2 Analysis of training data and derivation of decision rules*

The script examines each data point for certain aspects that help to assign structural information to the data point.

When looking at the data points, it is noticeable that in individual cases incorrect and thus illogical values occur during the historization. In line 5338 of figure 7 such an illogical recording occurs when the value for the currently measured outdoor temperature for a short period of time is given as 151.5 °C.

| | | |
|---|---|---|
| 5333 | 08.07.2021 10:02 | 20,73830032 |
| 5334 | 08.07.2021 10:14 | 20,93840027 |
| 5335 | 08.07.2021 10:34 | 21,13909912 |
| 5336 | 08.07.2021 10:41 | 24,31769943 |
| 5337 | 08.07.2021 10:41 | 21,2635994 |
| 5338 | 08.07.2021 10:42 | 151,5 |
| 5339 | 08.07.2021 10:42 | 21,3029995 |
| 5340 | 08.07.2021 10:58 | 21,50379944 |

**Fig. 7 -** System-related incorrect recording of a data point representing the outdoor temperature.

Due to such few erroneous recordings of the data points, the following rules are determined as percentages.

In the following, three exemplary functions are described which examine the logged values of a data point and classify this data point in a list on the basis of the trend.

### *Binary or analog:*

The script finds out whether a data point describes a binary or analog value by checking whether a data point has assumed 0, 1 or another value over the entire period under consideration. If the value is exclusively 0 or 1 over 99% of the entire period, it is a binary data point. If the data point also takes other values, it is an analog data point. Based on this rule, the data points are placed in the corresponding list for binary or analog data points.

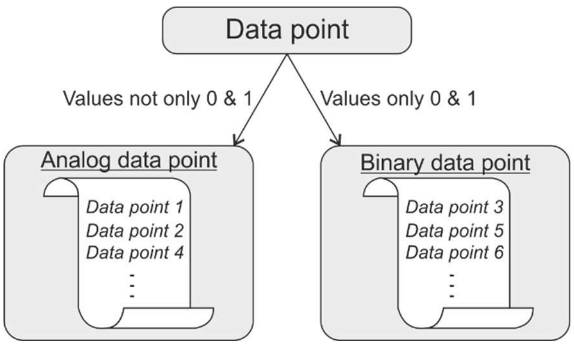


**Fig. 8 -** Binary or analog data point classification.

Based on this information, data points can be presorted in order to establish initial correlations in the further course. Binary data points rather represent operating messages such as pump switching commands to which analog values such as temperatures or positioning commands can react.

(The data used was provided with the help of the BACnet communication protocol. In BACnet, it is mandatory to specify whether a data point describes an analog or binary state. Therefore, it would not be necessary to determine this information with the help of a script. However, the approach taken in this paper is to consider data points independently of technology and to work only with the values provided and the associated timestamps).

### *Values over/under 100:*

When checking whether a data point takes analog values above or below the value 100, the information taken before can already be used as support. For this information only, data points need to be considered which were sorted into the list for analog data points by the function *'Binary or analog'* described before.

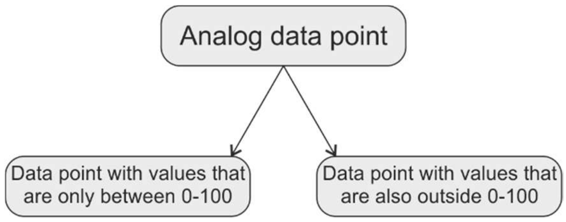


**Fig. 9 -** Data points with values that are only between 0-100 classification.

This information helps in the further course to give prognoses whether it is a data point which describes a valve position (0-100%). If the data point has values above 100 or below 0, it is most likely not a valve position.

### *Set points:*

A measured value behaves differently from than an associated setpoint. A setpoint changes suddenly and abruptly in contrast to a measured temperature value.

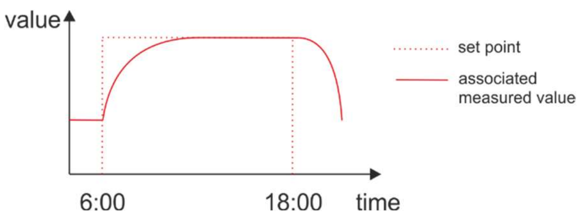


**Fig. 10 -** Qualitative representation of a setpoint and an associated measured value.

The data points were examined in terms of how often a data point executes a jump. A jump was declared as a 20% change of the value in a time window of less than one minute. As a percentage threshold value, it was found that every data point that makes these jumps (i.e. > 0%) can be declared as a control signal. Thus 18 of 22 setpoints were correctly recognized with two incorrect assignments. Thus, the data points were divided into "jumping" and "not jumping" data points.

***Room temperature:***

To check whether a data point could be a room temperature, only data points that have already been classified as analog data points are considered. Then, for each data point, the percentage of all recorded values that are within a typical temperature range for room temperatures is checked. This temperature range was set to the range 17 - 27. Applying this rule to the training data, it was found that if 50% or more of the recorded values are within this range, all existing room temperatures are correctly assigned.


**Fig. 11 -** Evaluation for the classification of a room temperature.

In Fig. 11 it is to be seen that many data points (red bars which go beyond the boundary line of 50%) which do not represent room temperatures also fulfill the rule and thus would be classified as room temperature. Therefore, it is necessary to clean up the lists after the first classification with logical regularities. For example, if a data point was declared as a potential room temperature and as a potential outdoor temperature, the outdoor temperature can be removed from the list of potential room temperatures afterwards. For example, an outdoor temperature could be recognized more precisely by the typical day & night cycle, which is more pronounced than for a room temperature, and thus be removed from the list of room temperatures.

***Heating temperature:***

Another list in which data points can be sorted is a list containing all data points representing heating temperatures. For the time being, no distinction is made between setpoints / measured values or flow/return temperatures. The decisive criterion here is again a range of values which in this case was set to 25 to 65. As a percentage threshold value, whether a data point represents a heating temperature or not, 55% was chosen.

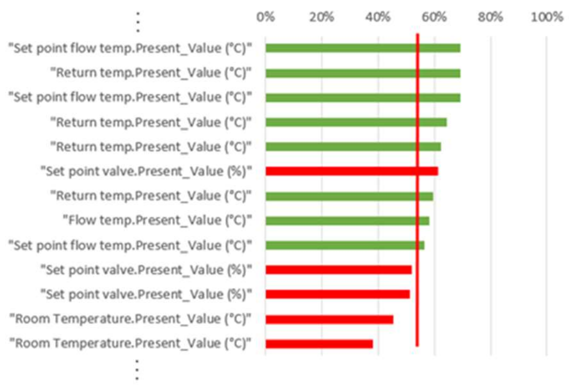


**Fig. 12 -** Evaluation for the classification of a heating temperature.

For the training data, this threshold detected all heating temperatures and assigned them to the list.

***Outside temperature:***

To detect outdoor temperatures, the percentage value was determined of how often the value of a data point corresponded to the historical monthly mean values (± 5 Kelvin) for the corresponding region (Greater Cologne-Germany). For the training data, a limit value of 40% resulted here.

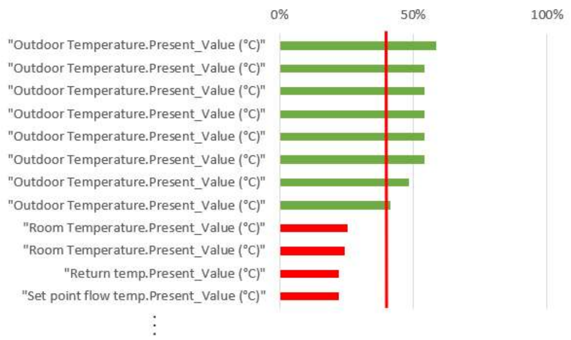


**Fig. 13 -** Evaluation for the classification of an outdoor temperature.

In this function, unlike the three rules presented previously, no incorrect data points were added to the list.

## *2.2 Finding relationships between data points*

Data points in building technology often interact with each other in a generally known way. For example, if a valve position changes or the status of a pump changes from "off" to "on", this is often accompanied by a time-delayed change in temperature. Fig. 14 shows such a relationship between flow temperature, valve position and pump status.

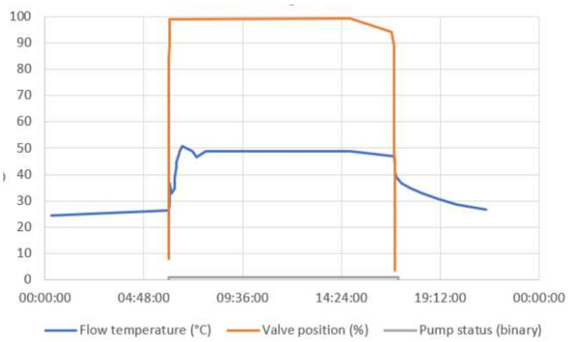


**Fig. 14 -** Relationship between flow temperature, valve position and pump status.

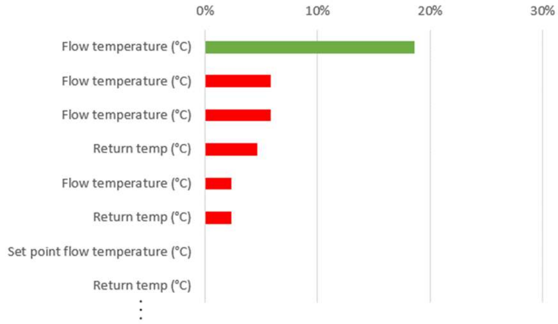


**Fig. 15 -** Evaluation for the relationship between a heating temperature and a pump status.

The data points were divided into individual lists using the rules applied in the previously described. These lists are used in the next steps to derive relationships between the individual data points, as explained in the previous section.

***Temperature-pump dependencies:***

To find out which of the temperatures is related to which pump switching command, the data points from the list for binary data points were first searched for state changes. If a state change was found in one of the data points, the timestamp of the state change was selected as a new search criterion. With this timestamp, each data point in the list 'Heating temperatures' was checked to see if there were values for this point in time that had changed significantly, but with a time delay. As a criterion how much a value must change, 10 Kelvin was defined. The time delay was chosen with 0-10 minutes.

The following graphic shows an exemplary result for a selected binary data point. By applying the rule described above, the correct data point (green bar) could be found. Compared to the other data points, it is by far the one that most frequently fulfills the search criterion.

In 19% of the cases, this supply temperature met the condition of a change of 10 Kelvin within 0-10 minutes after a change of state of the binary data point. The low percentage value results from inconsistent training data. The historicized trend data for heating circuits do not have the same time records for all data points. In the training data it happens that binary pump switching commands are available for a period for which the corresponding temperatures of the heating circuit have not been historized. For this reason, it was decided not to declare a percentage as the limit for this rule, but to take the data point with the highest percentage as the decision criterion.
Further dependencies that are examined between the data points are, for example:

- How often a data point declared as temperature follows a data point declared as valve position.
- How often a data point declared as temperature or valve position follows a setpoint.

## 4. Validation

In order to validate the previously derived rules, the rules were applied to historized data from another building technology system. The data points of this system relate to a ventilation system with supply and exhaust air fans, heat recovery and a heater including the associated heating circuit. The data points were historicized over a period of one week in November.

**Table 1** - Validation of the established rules based on data points of an air conditioning system. Cells with a green background mean that the respective data point has met the conditions of the respective rule.

| Datapoint | Is binary | Is analog | values over 100 | values under 0 | Roomtemp. | Heatingtemp. | Percentage of jumps | Outsidetemp. |
|---|---|---|---|---|---|---|---|---|
| Exhaust air temperature ventilation (°C) | 0% | 100% | 0% | 0% | 76% | 0% | 0,00% | 75% |
| Supply air temperature ventilation (°C) | 0% | 100% | 0% | 0% | 78% | 2% | 0,00% | 62% |
| Outdoor air temperature ventilation (°C) | 0% | 100% | 0% | 0% | 1% | 0% | 0,00% | 85% |
| Control signal heat recovery ventilation (%) | 63% | 37% | 0% | 0% | 0% | 3% | 0,23% | 0% |
| Control signal heating valve ventilation (%) | 22% | 78% | 0% | 0% | 7% | 33% | 0,08% | 8% |
| Set point supply air temperature ventilation (°C) | 0% | 100% | 0% | 0% | 100% | 0% | 0,00% | 36% |
| Control signal supply air fan ventilation (%) | 38% | 62% | 0% | 0% | 3% | 34% | 0,07% | 3% |
| Control signal exhaust air fan ventilation (%) | 40% | 60% | 0% | 0% | 3% | 37% | 0,07% | 2% |
| Operating signal pump heater ventilation | 100% | 0% | - | - | - | - | - | - |
| Operating signal supply air fan ventilation | 100% | 0% | - | - | - | - | - | - |
| Operating signal exhaust air fan ventilation | 100% | 0% | - | - | - | - | - | - |
| Return temperature heater ventilation (°C) | 0% | 100% | 0% | 0% | 19% | 74% | 0,00% | 19% |

- ***Binary or analog:***
  The rules that are supposed to distinguish whether a data point is a binary or analog data point work without error for this example. All three binary operating messages have been classified as such. A data point that has been classified as binary is no longer checked with the other rules.

- ***Values over/under 100:***
  The rules 'values over 100' and 'values under 100' should be used to distinguish data points that can take high/low numerical values (e.g. counters or high electrical power consumption) from data points that operate between 0-100 (e.g. system temperatures or valve positions). According to this rule, the historized data points of the ventilation system showed that each analog data point can be e.g. a system temperature or valve position and no data point could be excluded for the examination of the further rules.

- ***Room temperature:***
  The rule that checks a data point for a potential room temperature has detected three data points. The three detected data points are all related to the room temperature, but are not explicitly declared as such. Particularly From this example, potential for improvement of this rule can be derived. The data point "Setpoint room temperature" was declared 100% as potential room temperature, because it did not change over the entire period and was constantly in the defined number range. An addition to the rule stating that the data point must change by a defined value within a certain time period could act as an exclusion criterion here.

- ***Heating temperature:***
  The rule for detecting heating temperatures found the correct data point during validation. In this plant there was only one return temperature.

- ***Set points:***
  The recognition of the control signals also worked without errors. All four control signals were declared as such.

- ***Outside temperature:***
  The declaration of the outdoor temperature also worked well. Even though three data points were declared as potential outdoor temperature, the correct outdoor temperature has the highest percentage. This can also be improved by defining the rule in such a way that only the data point with the highest percentage is recognized as outdoor temperature.

**Table 2** - Relationship between binary datapoints and heating temperatures.

| Binary datapoint | Return temperature heater ventilation (°C) |
|---|---|
| Operating signal pump heater ventilation | 0% |
| Operating signal supply air fan ventilation | 0% |
| Operating signal exhaust air fan ventilation | 0% |

- ***Temperature-pump dependencies:***
  Table 2 shows the percentage values of the rule that is to find out which temperature follows which binary switching command. The only detected heating temperature could not be assigned to any of the three binary switching commands. The rule searches for a changing temperature after there has been a change of state of a binary signal. Figure 16 shows the recorded course of the heating temperature, the corresponding valve position and the operating message of the pump over a period of 24 hours. As can be seen, the control of the system is selected in such a way that the return temperature is controlled only by the valve position. The binary switching command of the pump is constant 1, so a dependence of the temperature to the pump cannot be found.

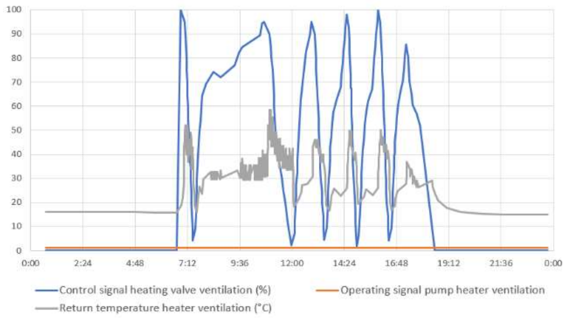


**Fig. 16 -** Relationship between return temperature, control signal valve position and pump status

But knowing that there is a heating temperature that follows a valve position and no associated pump switching command, predictions can be made about system behavior.

## 5. Problems and approaches

In the presented methodology, different problems due to inconsistent and incomplete data were identified. The goal must be to historize all data points of a building services system for a certain period of time. When recording a building's behavior, it is then important that the building's systems also operate accordingly. If valves do not move (e.g. because it is summer and no heating load is required), it is difficult to identify dependencies between data points.

## 6. Outlook and summary

This paper describes a first attempt to draw conclusions about the structural information of a data point from historized HVAC data points based only on the value and the corresponding timestamp.

Of course, HVAC-systems consists of larger and more complex systems than the systems taken as a basis in this paper. Nevertheless, with such simple rules, and knowledge of the known regularities of HVAC-systems, there is the possibility to generate information about the structure of a system.

The next steps in extending this method will be to further specify the rules for heating circuits and to define rules for other trades and components. The historization of the data can be done by a microcontroller which searches for the common communication protocols in a building automation network and historizes all data points in fixed intervals. These data can be used again as validation.

The rules described in chapter 3 can serve as a basis for one of the most widely used AI algorithms 'Decision Trees'. In 'Decision Trees', simple if-else relationships for data are also established and then interrelated using stochastic distribution in such a way that predictions can be made for data sets [10]. Whether the methodology presented in this paper can be used to automatically link HVAC data points with TMon applications could not be finally answered. In chapter 2 different methods were already presented which deal with similar questions. It will be interesting to investigate how far the methods can complement and support each other.

## References


[1] Facility management - Technical monitoring of buildings and building services, 6041, Verein deutscher Ingenieure, 10772 Berlin, Jul. 2017.

[2] H. Merz, T. Hansemann, and C. Hübner, Building automation: Communication systems with EIB/KNX, LON and BACnet. München: Fachbuchverl. Leipzig im Carl Hanser Verl., 2007.

[3] J. Granderson and G. Lin, "Building energy information systems: synthesis of costs, savings, and best-practice uses," Energy Efficiency, vol. 9, no. 6, pp. 1369–1384, 2016, doi: 10.1007/s12053-016-9428-9.

[4] M. Both, Maisch Nicolai, and Kämper Björn Prof. Dr. Müller, Jochen, "Automated performance monitoring of HVAC components by artificial intelligence," Rotterdam, May. 2022.

[5] A. Cartus, M. Both, and J. Prof. Dr. Müller, "Interoperability of semantically heterogeneous digital twins through Natural Language Processing methods," Rotterdam, May. 2022.

[6] John Petze, Unlocking the Value of Equipment System Data: Introduction to Project Haystack and Data Modeling for Equipment systems. [Online]. Available: https://project-haystack.org/download (accessed: Jan. 10 2022).

[7] B. I. John Petze, ASHRAE'S BACnet committee, Project Haystack and Brick schema collaborating to provide unified data semantic modelling solution [Online]. Available: https://www.ashrae.org/about/news/2018/ashrae-s-bacnet-committee-project-haystack-and-brick-schema-collaborating-to-provide-unified-data-semantic-modeling-solution (accessed: Jan. 10 2022).

[8] Sakshi Mishra, Andrew Glaws, Dylan Cutler, Stephen Frank, "Unified Architecture for Data-Driven Metadata Tagging of Building Automation Systems," National Renewable Energy Laboratory, Golden, USA.

[9] H. R. Kranz, BACnet Building Automation 1.12: Basics in German language, 3rd ed. Karlsruhe, Baden: cci-Dialog, 2013.

[10] D. N. Sidorov, Ed., Machine Learning for Energy Systems. Basel, Switzerland: MDPI - Multidisciplinary Digital Publishing Institute, 2020. [Online]. Available: https://directory.doabooks.org/handle/20.500.12854/69401