



UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

DIPARTIMENTO DI

Corso di Laurea Magistrale

in

Contents

1	Semantic Web	3
1.1	Introduction to the Semantic Web	3
1.1.1	Linked data	4
1.1.2	Vocabularies	6
1.1.3	Query	7
1.2	Ontologies for smart buildings	8
1.2.1	Brick	8
1.2.2	Semantic Sensor Network ontology	9
2	Model creation	12
2.1	State of the art	12
2.2	IBM Research approach	13
2.2.1	Semantic Mapping	15
2.2.2	Physic model inference	17
2.3	Example of the whole approach	19
3	Diagnosis Algorithm	26
4	IoT Brain	27
4.1	IoT BRAIN and KITT	27
4.1.1	Graph database vs. RDF Store	28

Introduction

Industrialised countries are facing challenges that are strictly related to energy consumption. Although great efforts in studying novel energy sources are profused by both public and private entities, these solutions need to be sustained by efforts aimed at the reduction of energy wasting. Around 32% of total energy consumption in industrialized countries is used for electricity, heating, ventilation and air-conditioning (HVAC) in commercial buildings. Studies have shown that early detection of faults in those systems and their operation could lead to energy savings between 15% to 32%, improving the overall buildings efficiency. This scenario does not only offer great opportunities related to environmental protection but gives companies a chance for sensible expense reduction. Around 14% of commercial buildings in the US (as of 2012) deployed Building Management System (BMS), supported by the increasing number of available sensors and improvement in the Internet of Things technologies. These efforts end up facing scalability issues due to the lack of common schema for data coming from a variety of building, vendor and location specific sources. Applications developed for ...

Chapter 1

Semantic Web

In this chapter will follow a brief description of the concept of ontology and of Semantic Web as defined by W3C. This chapter will then further detail the ontologies and metadata schemata used in the thesis project, namely Semantic Sensor Network (SSN) and Brick. It is to note that the work presented in this thesis is not directly using all of these technologies, but it draws from their key concepts to reach some goals. It is therefore worth to mention such concepts here.

1.1 Introduction to the Semantic Web

Semantic Web is a term that identify an evolution of the web in which every resource available through the web is associated with a semantic meaning. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Semantic web layers are defined by W3C as:

- linked data
- vocabularies
- query
- inference

1.1.1 Linked data

Linked Data lies at the heart of what Semantic Web is all about: large scale integration of, and reasoning on, data on the Web. To make the Web of Data a reality, it is important to have the huge amount of data on the Web available in a standard format, reachable and manageable by Semantic Web tools. Furthermore, not only does the Semantic Web need access to data, but relationships among data should be made available, too, to create a Web of Data (as opposed to a sheer collection of datasets). This collection of interrelated datasets on the Web can also be referred to as Linked Data.

Resource Description Framework

RDF [7] is a description language, which aims to capture the semantics of data represented on the Web. The RDF standard is based on several fundamental concepts. The first such concept is “resource”. Resources are the basic objects, or “things”, that are to be described in the domain (e.g. rooms, floors, sensors etc.). All resources are identified through a unique, global identifier called the universal resource identifier (URI). In most applications, the URI is the uniform resource locator (URL) of a Web page, a part of a Web page, or a link to a document available on a Web server. However, the URI is a more general concept, the only condition is that it uniquely identifies a resource. Examples of resources could be “unibs:florenzi.002” as well as “unibs:thesis1234”, or “https://www.w3.org/RDF/”. Another essential concept is “property”. Properties describe relations between resources and following the previous example we could say that “has_author”, “has_name”, “has_subject” are properties. Having defined both resources and properties, statements can be derived. Statements, also known as triples, in RDF have the basic form: subject-predicate-object, where the subject is an RDF resource, the predicate is a RDF property, and the object is another RDF resource or a literal (a name, a number, code, etc.). Still referring to the example, the statements could be “unibs:thesis1234 has_author unibs:florenzi.002”, “unibs:thesis1234 has_subject https://www.w3.org/RDF” and “unibs:florenzi.002 has_name “Fabio Lorenzi””. RDF data models end up being represented as graphs where subjects and objects are nodes while properties are the edges. The resulting

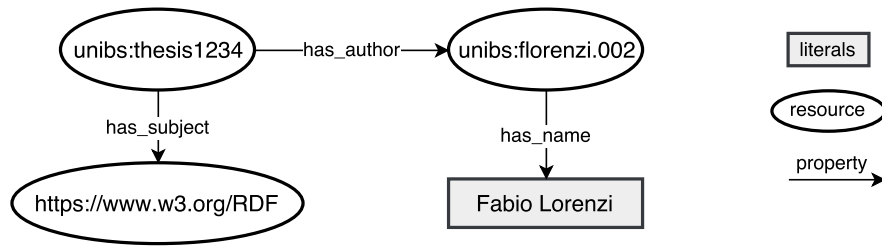


Figure 1.1: Graph representation of the example RDF model

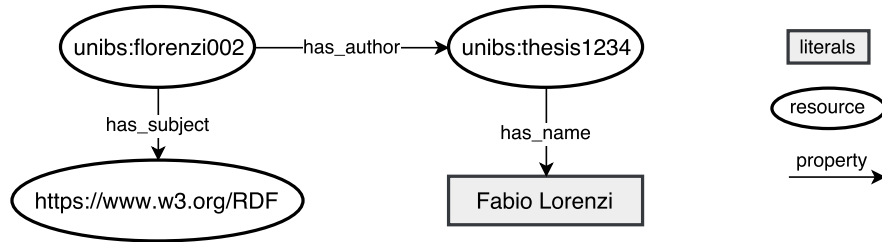


Figure 1.2: Correct RDF model with wrong semantic

graph, given the RDF notation, in the example is shown in Figure 1.1.

It is to note that RDF is a conceptual model for representing linked data, but as far as it goes, no semantic is implied in this representation, looking at this model is impossible to understand the meaning of the resources or property, nor to understand if those elements reflect and adhere to rules of a particular domain. In the example above there is no explicit information telling us that the resource “unibs:florenzi.002” is a person and that “the unibs:thesis1234” is a book. Furthermore the RDF model represented with “unibs:florenzi.002 has_author unibs:thesis1234”, “unibs:florenzi.002 has_subject `https://www.w3.org/RDF`” and “unibs:thesis1234 has_name “Fabio Lorenzi””, yielding the graph representation in Figure 1.2, is perfectly fine, although not semantically correct¹. Adding semantic to an RDF model, that is detailing the domain ontology, is left to other frameworks as explained in **subsection 1.1.2**.

¹from a common sense point of view. No semantic is explicitly declared

1.1.2 Vocabularies

On the Semantic Web, vocabularies define the concepts and relationships (also referred to as “terms”) used to describe and represent an area of concern. Vocabularies are used to classify the terms that can be used in a particular application, characterize possible relationships, and define possible constraints on using those terms. There is no clear division between what is referred to as “vocabularies” and “ontologies”. The trend is to use the word “ontology” for more complex, and possibly quite formal collection of terms, whereas “vocabulary” is used when such strict formalism is not necessarily used or only in a very loose sense. Vocabularies are the basic building blocks for inference techniques on the Semantic Web.

Resource Description Framework Schema

RDFS defines a set of RDF resources useful for the description of other RDF resources and properties. Through RDFS it’s possible to define and detail the semantic of a RDF model. Among the various concepts introduced with RDFS there are

- `rdfs:Resource`: everything that is described in RDF is a resources
- `rdfs:Literal`: it is text string
- `rdfs:Property`: it represent the properties
- `rdfs:Class`: it is similar to the concept of type or class in an object oriented programming language.
- `rdfs:subClassOf`: it specify inheritance, eventually multiple, between classes.
- `rdfs:range`: it tells which resources are permitted as object in a statement
- `rdfs:domain`: it tells which resources are permitted as subject in a statement

it is to note that the RDFS is a RDF model itself and it’s used for explicitly declare the domain knowledge of a given RDF model. We refer to the RDFS as an ontology and to the RDF model as the instances of such ontology. Following

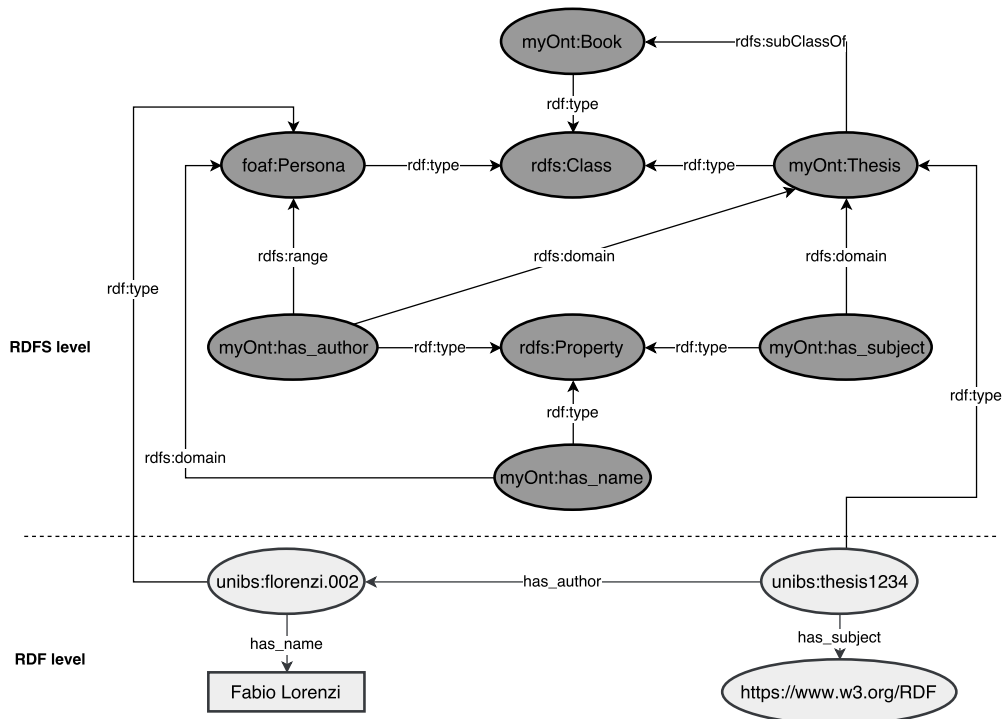


Figure 1.3: RDFS representation of an ontology

RDFS, it is possible to model domain knowledge for the example in **section 1.1.1** as in Figure 1.3.

This “layering” approach is what allows for the design of reusable ontologies and gives opportunity to automatically infer new knowledge based on the RDFS model of the data. Still looking back at the example, let's suppose that the statement “unibs:florenzi.002 rdfs:type foaf:Person” wasn't in the ground knowledge of the RDF model. Even without this explicit information wasn't available, a reasoner would have been able to infer that statement (dotted line) given the semantic description of the property “myOnt:has_author” (bold line). The process is represented in Figure 1.4.

1.1.3 Query

Just as relational databases or XML need specific query languages (SQL and XQuery), the Web of Data, typically represented using RDF as a data format, needs its own, RDF-specific query language and facilities.

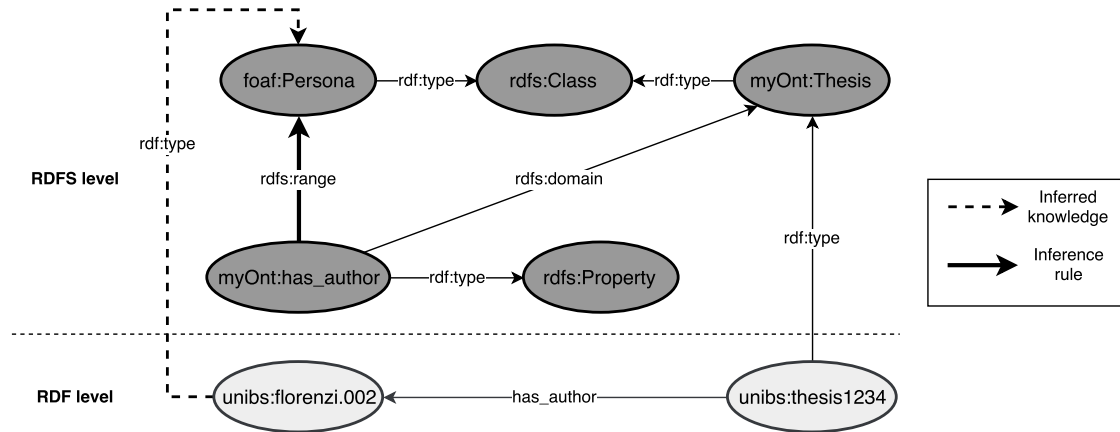


Figure 1.4: Inference process of new knowledge

SPARQL Protocol and RDF Query Language

A mean for querying RDF models is defined by SPARQL [10] query language. Technically, SPARQL queries are based on (triple) patterns. As RDF can be seen as a set of relationships among resources, SPARQL queries provide one or more patterns against such relationships. These triple patterns are similar to RDF triples, except that one or more of the constituent resource references are variables. A SPARQL engine would return the resources for all triples that match these patterns. To represent a pattern (or query) to run against the data model, SPARQL defines a series of keywords

1.2 Ontologies for smart buildings

1.2.1 Brick

Brick schema [1] defines domain specific concepts aimed at describing sensors in a building with its context. It is composed of three parts

- a class hierarchy of entities describing various building subsystems
- a minimal set of relationships for connecting entities
- a methods for encapsulation in the form of Function Blocks

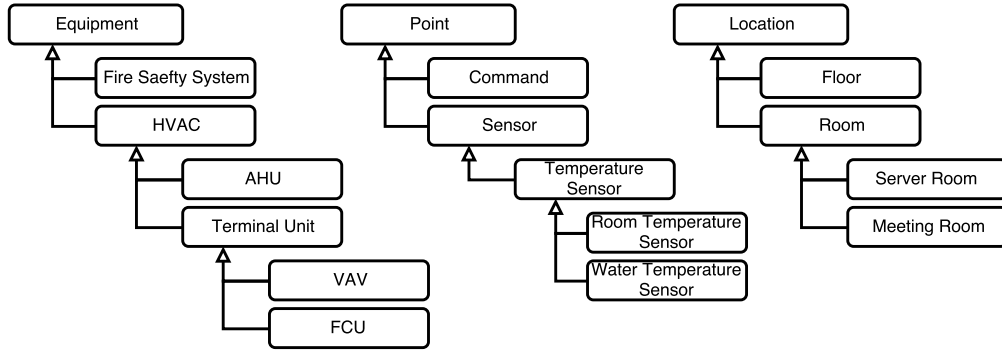


Figure 1.5: Subset of Brick hierarchy

the main concept in Brick is the concept of tag introduced with project Haystack [6], which is further enriched by an ontology that cristallizes the concepts defined by tags. A tag is a flexible framework for annotating metadata to building data points. In Brick tags are grouped together as sets, named tagset, to represent entities; the tags room, temperature and sensor can be grouped together as room temperature sensor representing a single entity. The concept of tagset is well integrated with the class hierarchy concept of the ontologies: a room temperature sensor is a subclass of a generic temperature sensor. As the main goal of Brick is to represent points in a building’s BMS, Points is one of the main classes of Brick, which subclasses define specific type of points like Sensor. The most common concepts a Point can be related to are Location, Equipement and Measurement so those classes, together with the Point class, form the core classes of Brick. Those classes are further subclassed to form a hierarchy. A subset of the hierarchy is shown, as example, in Figure 1.5. able to comprehend every entity of a building. Alongside classes, relationships play a crucial role in connecting entities and thus providing a context for many applications. Brick designed a set of minimal, intuitive and multipurpose relationship fundamental in capturing existing relationship. Figure 1.6 shows the Brick core concepts and the designed set of relationships.

1.2.2 Semantic Sensor Network ontology

While Brick schema was used to model the building domain, the Semantic Sensor Network (SSN) ontology [2] was chosen because it provides a different view

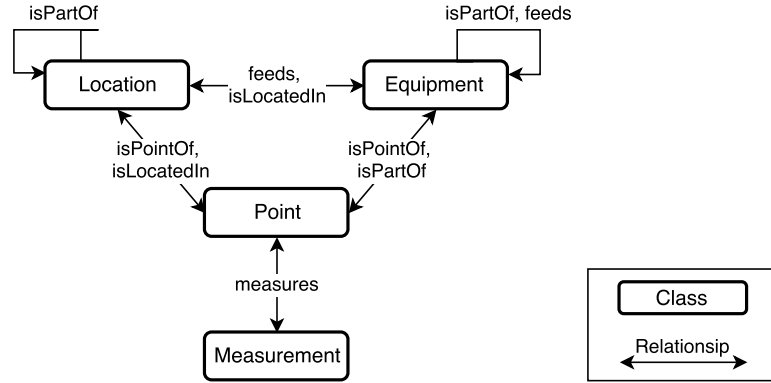


Figure 1.6: Brick schema core concepts

of the various aspect of the sensing process, beyond the specific building domain. Peculiar to the SSN Ontology (SSNO) is the Stimulus-Sensor-Observation design pattern. This pattern differentiate between the concept of `ssn:Sensor` for physical devices that take measurements in form of `ssn:Observation` and `ssn:Stimulus` that is the actual change in the enviroment that caused a particular observation, so that it possible to differentiate that stimuli occur in the reality even if no sensor is monitoring them, thus modeling unobservabel states of a system, and that observation made from a sensor can be different from the real event (e.g errors, failures, noise). These concepts aside, SSN provide other useful classes that are `ssn:FeatureOfInterest` that represent generic entities which status is to be monitored, like rooms, fans or floors and `ssn:Property` that embodies specific physical quantities that a sensor ought to monitor. All the classes are connected by a set of relationships that are `ssn:hasProperty` that connects a `ssn:FeatureOfInterest` with meaningful physical properties, like a temperature in a room; `ssn:isProxyFor` represent that a given `ssn:Stimulus` influence a `ssn:Property` as in “another occupant in a room raises its temperature”; `ssn:includeEvent` create a dependency between an observation and a event that causes it; `ssn:observedBy` links an observation to the sensor that could measure it. Figure 1.7 show the resulting SSN ontology.

Extending SSNO

As far as it goes Brick and SSN are good ontologies for modeling a building and its sensor network, but in a Energy Management System (EMS) context, and for

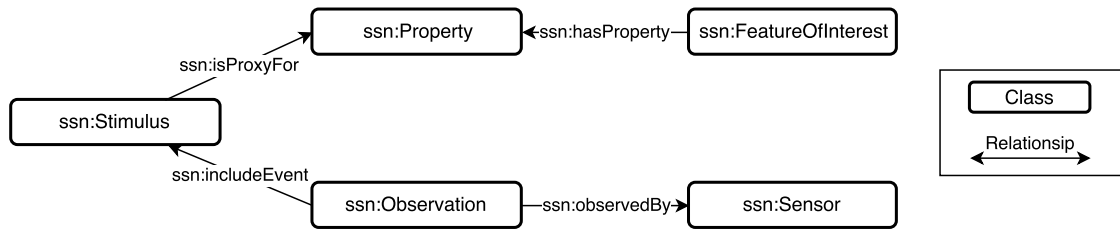


Figure 1.7: Semantic Sensor Network (simplified) ontology

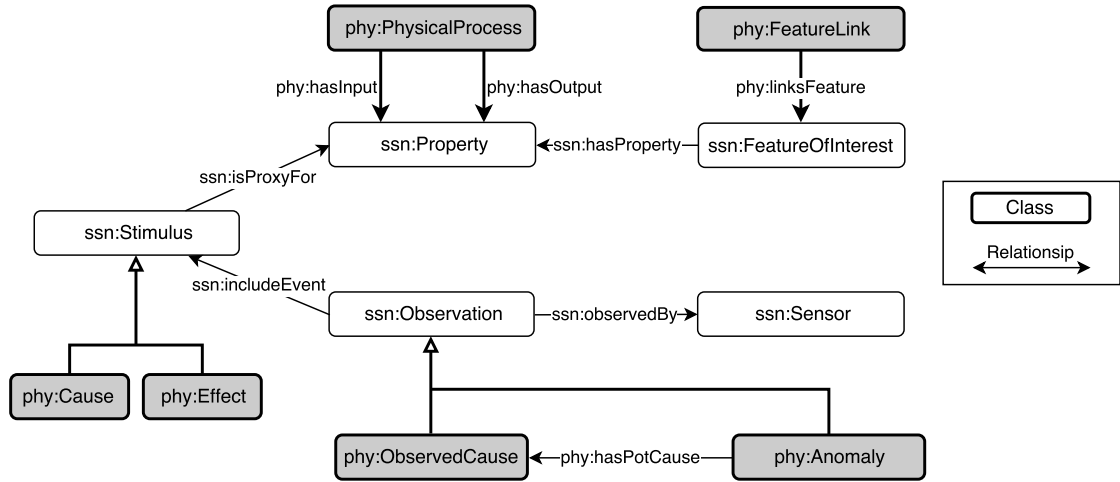


Figure 1.8: Extended SSN ontology. Extensions are highlighted with bold line and a light grey background.

FDD applications in particular, it is essential to understand the cause-effect relationships in the building. In order to do that, IBM Research extended the SSNO through the introduction of additional classes and references as in Figure 1.8. `phy:PhysicalProcess` represents laws of physics through a simplified model. Different subclasses of physical processes are derived from LTI system processes and will be described in chapter 2. `phy:FeatureLink` tells which kind of relationship exist between two `ssn:FeatureOfInterest`, such a link could be a wall between two rooms or a window between a room and the outside. `phy:Cause` and `phy:Effect` are subclasses of `ssn:Stimulus` useful for differentiate whether said stimulus is a cause or an effect of a `ssn:Observation`. `phy:Anomaly` is a subclass of `ssn:Observation`, it describes out-of-range observations that need to be diagnosed, like a high temperature in a room. `phy:ObservedCause` is still a subclass of `ssn:Observation`; it is used for describing a measurement that can be a cause of an anomaly.

Chapter 2

Model creation

This chapter will outline the approach suggested by IBM Reserch for tackling the challange introduced in **Introduction** regarding fault detection and diagnosis (FDD).

2.1 State of the art

Nowadays, as extensively analyzed by Katipamula and Brambley [4], FDD approaches fall in either one of the following three categories: physycal model based approaches, data driven approaches and rule-based approaches. The rationales behind these approaches are detailed in section 2.1, section 2.1 and in section 2.1. Still, each of these approaches share common limitations that can be summarized as

- they can be applied to a single, specific building
- they requires large manual effort and expertise
- deploy can take several weeks
- they neglect strong interactions between systems

Physycal model approach These approaches require a physycal model (e.g ordinary differential equations) of the building and its components. They are

highly precise in diagnosing faults given a correct model, however, deriving the models is no trivial task and requires times and expertise. On top of this, derived models are building, system and location specific and are difficult to adapt to other buildings than the one they are thought for even if they share similar structure and similar components, thus limiting the scalability of this approach.

Data driven approach Data driven approaches completely rely on building's sensor data. They assume that access to a large dataset of hystorical data is granted. There's little to no need for any *a priori* knowledge of the processes involved. Black-box data driven approaches derive the model in the form of a input-output relationship which parameters aren't correlated with the actual physical parameter (e.g artificial neural networks, regression) while grey-box data driven approaches take advantage of simplified physical relationships between measured quantities (e.g principal component analysys) and rely on statistical methods for estimating their parameters. Even though these methods don't suffer from scalability issues they are limited to fault detection and lack in diagnosis capabilities.

Rule-based approach The rule-based approach is the most common in traditional FDD applications. It uses domain knowledge and expertise in order to derive a series of simple *if-then-else* rules or some kind of decision trees along with a series of thresholds and confidence intervals; during system operations data are evaluated against this rules and countermeasures are eventually taken. Even though this is the most common approach, it still needs a lot of manual effort and requires access to domain knowledge so these requirements still limit portability and scalability of applications based on this technique.

2.2 IBM Research approach

The approach developed by IBM Research is based on a combination of the three aforementioned concepts to overcome their disadvantages. It models high level physical processes in the systems to derive diagnosis rules, it parametrize the rules using data analytics techniques and applies them during systmes operations. The strength of the semantic approach lies in its capability to automatically infer knowl-

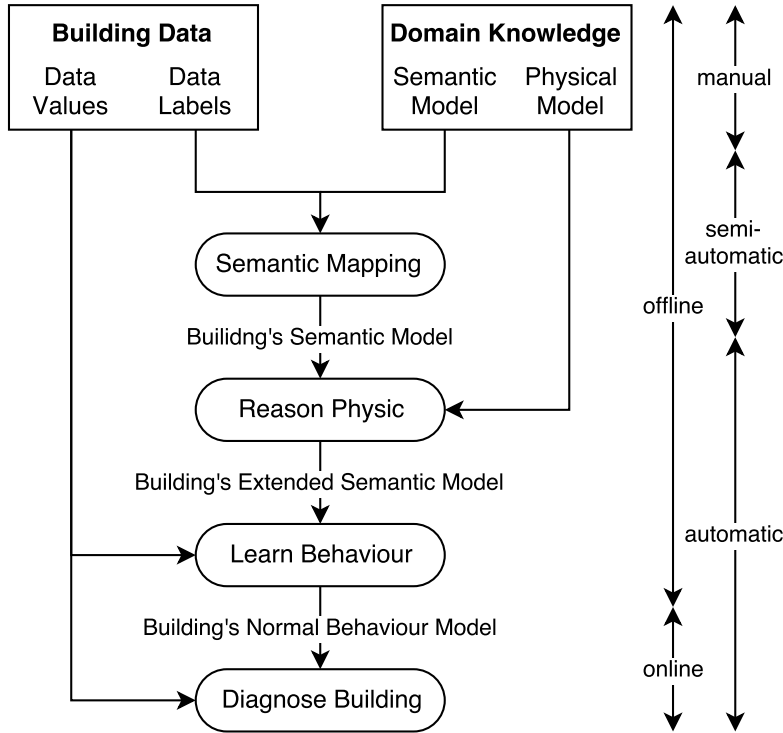


Figure 2.1: IBM Research approach, overview

edge given a building, its sensors and their data. This lead to a semi-automated process that limit the manual effort needed, as shown in Figure 2.1.

The proposed approach needs 3 inputs to produce the diagnosis of a given anomaly

- building data: assumed available in the building's Building Management System (BMS)
- semantic model: specified through the use of concepts from domain ontologies Brick (subsection 1.2.1) e SSNO (subsection 1.2.2)
- physycal model: the model of the physical dependencies between subsystems of the building. The physical processes are modelled through concepts presented in the extended SSNO (section 1.2.2)

these inputs are processed through the various stages to produce the complete

model of the building and its internal processes. Details about the different phases are given in the following pages.

2.2.1 Semantic Mapping

Building a semantic model for a specific building needs knowledge on the type of sensors, meters and various equipment available in said building, their location and their semantic meaning in the chosen ontologies. These informations are usually stored in BMS through labels that, despite some kind of standardization efforts, are usually vendor specific; in some cases even the same vendor ends up changing its labeling schemes throughout the years. Even though there is no common ground to evaluate these labels, usually they are comprised of a series of abbreviations, eventually separated by some special character (e.g “-”), that gives informations about the device’s ID, function and location (or the equipment it is part of). An air temperature sensor in a room on ground floor can be labeled AIR_TEMP_R3GF while another air temperature sensor in a room on the first floor may ends up being labeled as RATSF1R9; both the sensors represent a common semantic resource in an ontology and need to be mapped that concept. This process is called semantic mapping. Existing approaches are differentiated in three different types [9]

- semi-automatic: they offer a tool assisting the user in labelling the point to corresponding semantic types and are based on advanced text mining techniques (e.g regular expressions, classifiers)
- data driven: these approaches try to recover the meta-data given the time-series. It is based on the concept that different data-points exhibiting similar timeseries behaviour should be similar themselves.
- active feedback: in these kind of approaches a series of known events are injected into the system and datapoints semantic model is derived observing the effect of the event injection.

Building Energy Asset Discovery tool

BEAD [5] is the tool developed by IBM Research for assisting user in the process of discovery and tagging of sensors. The tool is based on dictionaries of the type

Table 2.1: Dictionary extended with assets' informations

Label	Asset	Marketset
U6_RAT	AHU6	{Return, Air, Temperature}
U6_DAT	AHU6	{Discharge, Air, Temperature}
AHU7_RAT	AHU7	{Return, Air, Temperature}
AHU7_SAT	AHU7	{Supply, Air, Temperature}
AU9_RET_TEMP	AHU9	{Return, Air, Temperature}
AU9_SUP_TEMP	AHU9	{Supply, Air, Temperature}

$\mathcal{D} : \mathcal{A} \rightarrow \mathcal{MS}$ where \mathcal{A} is the set of all relevant acronyms and \mathcal{MS} is the set of markerset; each markerset is a set of markers, or keywords. These concepts are closely related to those of tagset and tag found in the Brick ontology and this duality allows the association of the sensor with the correct semantic type as described by Brick (see 1.2.1). The dictionary contains the most common acronyms found in BMS and it is further extended by the markers (tags) from the Brick ontology, such that a tag maps to itself, e.g $\text{Temperature} \rightarrow \{\text{Temperature}\}$. The tool take a label from the BMS and compute a similarity score against the dictionary entries and guide the user in the labeling process. For example, given the following dictionary

- d1:** $\text{RAT} \rightarrow \{\text{Return, Air, Temperature}\}$
- d2:** $\text{RAT} \rightarrow \{\text{Room, Air, Temperature}\}$
- d3:** $\text{SAT} \rightarrow \{\text{Supply, Air, Temperature}\}$
- d4:** $\text{OAT} \rightarrow \{\text{Outdoor, Air, Temperature}\}$
- d5:** $\text{Temp} \rightarrow \{\text{Temperature}\}$
- d6:** $\text{Temperature} \rightarrow \{\text{Temperature}\}$

the label RATSF1R9 (Room Air Temperature Sensor, Floor 1, Room 9) has a high similarity with both $\text{RAT} \rightarrow \{\text{Return, Air, Temperature}\}$ and $\text{RAT} \rightarrow \{\text{Room, Air, Temperature}\}$, so it's up to the user to understand the meaning and choose the right markerset. In a similar fashion it is possible to extract information on the location of a sensor or the asset it is related to, as shown in Table 2.1.

this process clearly requires human supervision, hence the definition of semi-automatic, but through this approach the operator has to manually inspect a smaller subset of all the possible marketsets, thus taking minutes instead of several

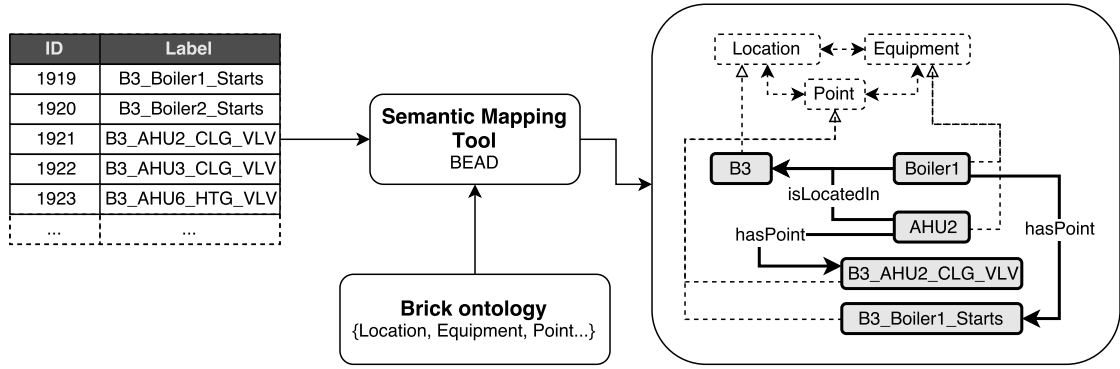


Figure 2.2: Semantic mapping process

weeks to complete the semantic mapping; practical experiments demonstrated that through this approach the decrease in number of labels to analyze is the 7.5% of the total number of BMS labels [9]. The output of this process is the semantic model of the specific building according to the chosen ontology. A flowchart illustrating this process is shown in Figure 2.2.

2.2.2 Physic model inference

The semantic model obtained from the semantic mapping phase is a model of the building architecture, assets, sensors and their location across the building. The model still lacks information about the physic processes taking place in the building and how the actual physic model is monitored and influenced by the points (sensors, setpoints etc.) in the building. Manually building the physical model of a building is unpractical and costly, thus the developed framework approaches this challenge through the use of semantic inference techniques. This phase requires as inputs the semantic model of the building and an appropriate ontology that models the physics variables. Given those inputs, through the reasoning process is possible to automatically add the correct dependencies and physical processes to the model. This process is separated in two phases:

- physical properties inference: it recreates the physical variables that are involved in physical processes. The properties are either mandatory or optional.

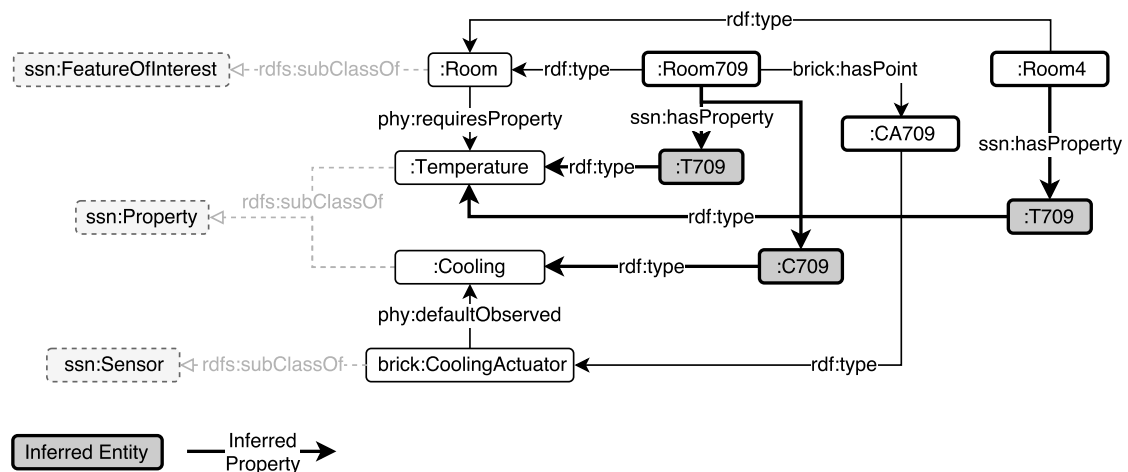


Figure 2.3: Process of inference of mandatory and optional properties.

- physical processes inference: once the physical properties are set in place, it is possible to link them together according to the domain knowledge specified in the upper ontology.

further informations about the concept of reasoning, reasoner and further details about implementation choices for this approach are given in ??

Physical properties inference Properties represent physical quantities in the real world and can be mandatory or optional. Mandatory property are characteristic of a `ssn:FeatureOfInterest` and need to be created for each instance of that feature. This behaviour is specified in the ontology of the physic through the use of the `phy:requiresProperty` relationship, so that during the reasoning mandatory properties can be automatically created. For example, in Figure 2.3 it is shown that the concept of `:Room` `phy:requiresProperty` `:Temperature`, that means that every instance of a `:Room` needs to be associated with a property that is an instance of `:Temperature`, or in plain english that every room has a temperature. Optional properties, on the other hand, need to be created only if some requirements are met, for example a `:Room` have a `:Cooling` property only if a `:CoolingActuator` is present in said room, and it is modeled in the ontology as `:CoolingActuator` `phy:defaultObserved` `:Cooling`. In Figure 2.3 it can be seen that only `:Room709` has a `:Cooling` property since it's the only room that has a `:CoolingActuator`. The

:requiresProperty and :defaultObserved properties¹ are concept level references and model general knowledge. General knowledge describe the domain without bothering about the ground truth. That means, for example, that the concept of :Room requires the existance of the concept of :Energy even if in the model of the real world there aren't any rooms. The existance of said properties is the key for the automatic reasoning process.

Physical processes inference Once all the actual properties have been infered for a specific configuration, it is time to append to the model informations about the physical processes involving such properties. As already stated in section 1.2.2, physical processes are modeled as Linear Time Invariant (LTI) systems based on the fact that Multiple Input Multiple Output (MIMO) processes can be decomposed into multiple Multiple Input Multiple Output (MISO) processes. A MISO process can be further subclassed as a Single Input Single Output process. Each process follows the rules

Rule 1 a process has at least one input and exactly one output

Rule 2 when a process has multiple inputs, they superpose additively

given these rules, it is possible to define a class hierarchy of LTI processes that are used to model a physical system. A subset of said taxonomy is available at Table 2.2.

2.3 Example of the whole approach

In order to better explain the concepts presented in the previous sections an example of the full process applied to a room, facing the outside environment, with a cooling Air Handling Unit (AHU). At the end of the section it is shown that a full model of the room, its internal and external processes is derived. This model will then be used during the explanation of the diagnosis process. Looking back at Figure 2.1, we start by defining the inputs needed for the process (sharp top boxes).

¹intended as the semantic concept of relationship between entities, not to be confused with domain knwoledge properties like :Temperature.

Table 2.2: Taxonomy of the semantic representation of LTI processes

Semantic concept	Parent	Formal definition
MISO	Process	$y_i(t) = \mathbf{c}_i \cdot \mathbf{x}_i + \mathbf{d}_i \cdot \mathbf{u}_i(t) = f(\mathbf{u}(t))$
SISO	MISO	$y_i(t) = \mathbf{c}_i \cdot \mathbf{x}_i + d_i \cdot u_i(t) = f(u(t))$
Positive Correlated (PC)	SISO	$y(t) \propto u(t)$
Negative Correlated (NC)	SISO	$y(t) \propto -u(t)$
Proportional (P)	SISO	$y(t) = k_p \cdot u(t)$
Positive Proportional (PP)	PC, P	$y(t) = k_p \cdot u(t), k_p \geq 0$
Negative Proportional (NP)	NC, P	$y(t) = k_p \cdot u(t), k_p < 0$
Negation (N)	NP	$y(t) = k_p \cdot u(t), k_p = -1$
Integral (I)	P	$y(t) = k_I \int u(t) dt$
Derivative (D)	P	$y(t) = k_D \cdot \dot{u}(t)$
Lag (PTn)	P	$\sum_{i=0}^n C_i \cdot T^i \cdot y^{(i)}(t) = k_{PT} \cdot u(t)$
1 st order lag (PT1)	PTn	$T \cdot \dot{y}(t) + y(t) = k_{PT} \cdot u(t)$
Delay (τ)	SISO	$y(t) = u(t - \tau), \tau > 0$
Multiplicative (M)	MISO	$y(t) = u_1(t) \cdot u_2(t)$

Building data and semantic model It is assumed that some kind of BMS is available in the building so that through the BMS it is possible to retrieve both labels (for semantic mapping) and data of various sensors (useful for diagnosis purposes, see). The room analyzed in the example will be named ROOM1. In ROOM1 an air temperature sensor and an occupancy sensor are available. Room1 is also equipped with an AHU, named AHU709 which internal components are observed by a fan mode sensor, a supply air flow sensor, coil water temperature sensor and a coil water flow sensor. This AHU709 is fed by a chiller named C1 equipped with a power sensor. Let the BMS contains the data as in Table 2.3 and let use Brick (1.2.1) as a domain ontology. through the semantic mapping tool (2.2.1) it is possible to map the labels in the BMS to the Semantic type in the brick ontology. The user can also give the systems additional information about the architecture of the building and the location of equipments and sensors. It is then possible to map :B3 as a brick:Location, :ROOM1 still as brick:Location, and then link one to the other as :ROOM1 brick:isPartOf :B3 and so on. The tool will also discover the semantic type of the sensor and will be able to create the correct instances, for example it will add to the model an air temperature sensor named :ATSROOM1 and will link it as a point to :ROOM1. The output semantic graph

Table 2.3: Example of BMS content

ID	Label
1234	CCPC1B3
1235	CWFU709
1236	CWTU709
1237	SAFAHU709C1
1238	FMAHU709
1239	OCCR1
1240	ATSROOM1
1241	ATS_Outside

can be seen in Figure 2.4 where the instances created from the BMS and their mutual relationships are bold, while the brick ontology concepts and relationships are light grey. This is the building semantic model, that will be used together with the physic model described in the next paragraph to infer knowledge about the physics properties and relationships in the building.

Physic model The physic model is a model of the physic phenomena occuring in an enviroment. It defines physical variable and processes and their interaction in the context of a specific domain. Creating a model of physic laws is a time consuming task that involve expertise from different fields, from engineering to physics, that are very domain specific. The good side of the semantic approach is that once the model is derived from a general point of view, it is automatically extended for every instance of a building, indipendently from its architecture. It is to note that the domain model for the physic does not necessarily model exact physical relationship but just the principal ones, since those are the most relevant to the approach, however this choice is purely paractical rather than a limitation of the approach itself and given sufficient time and knowledge it is possible to enrich the model in order to close the gap with the reality. Here will be presented a model for the physic of a room, equipped with a AHU fed by a chiller. The model is based on a simplification based on the assumption of fully mixed air, that allows for the modeling of a whole space as a single node with uniform temperature. Wetter [12] provide an insight of the physic involved in the example. It starts describing a

3. a room's temperature is influenced by the energy in a first order lag process [Process, PT1]
4. rooms can be cooled, if they have a cooling actuator [Property, optional]
5. cooling is negative proportional (NP) with the room's energy
6. room energy depends from the adjacents enviroment's temperature proportional to the shared surface A_i and the heat transfer coefficient h_i , hence the dependency is a positive proportional process [Process, PP]

further analyzing the physic involved, bring to the conclusion that

$$Q_{Occ} = q_{lat}n_{occ} \quad (2.5)$$

where q_{lat} is the latent heat gain and n_{occ} the number of occupants in the room. From Equation 2.2 and 2.5 it follows that

7. rooms have a number of occupants [Property, Mandatory]
8. a room's internal energy is positively correlated with the number of occupants [Process, PP]

additional knowledge can be added if the physics of the system is known; for example, it is possible to further enrich the model by specifying the internal processes that regulate a AHU. Wang et al. [11] provide a simplified model for modelling the cooling load of an AHU through Equation 2.6

$$Q_{Cool} = \frac{c_a \dot{m}_a^e c_c \dot{m}_c^e}{c_a \dot{m}_a^e + c_c \dot{m}_c^e} (\vartheta_a - \vartheta_c) \quad (2.6)$$

this, given that ϑ_a is the mixed air temperature, \dot{m}_a the air flow rate, ϑ_c the chilled water temperature and \dot{m}_c the chilled water flow rate, leads to the following conclusions

9. cooling is a complex function of the air temperature, the chilled water temperature, the supply air flow and the water flow in the coil [Process, MISO]

finally, Cui and Wang [3] provide a simple model of a chiller's physic, which equation can be expressed as

$$\dot{m}_c c_w (\vartheta_{cret} - \vartheta_c) = \eta_{COP} W \quad (2.7)$$

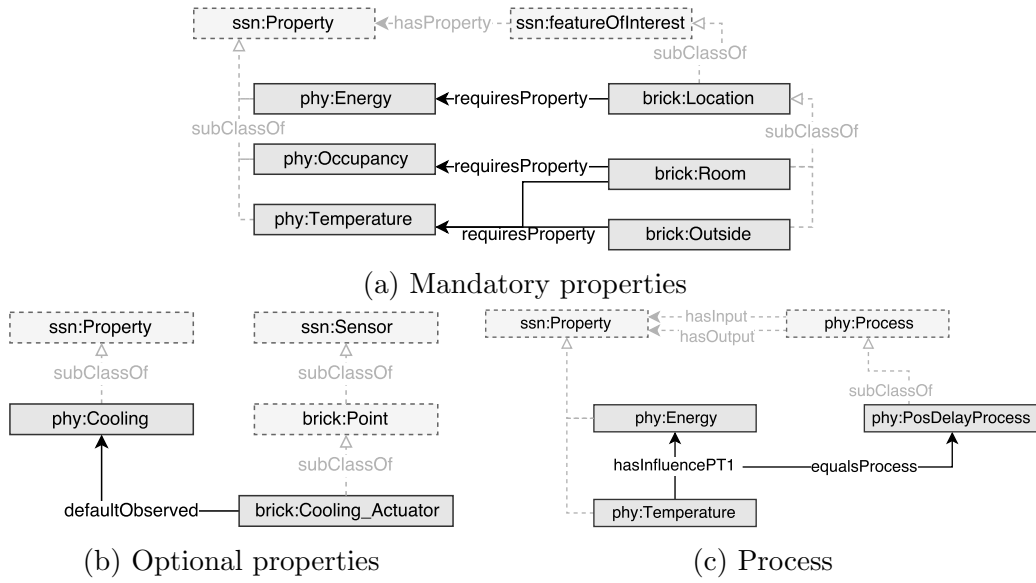


Figure 2.5: Patterns for definition of a model for a system's physic

from which it can be derived

1. the chilled water tempearture decrease with the increase of the power consumption [Process, NP]
2. the water flow rate increase as the power consumption of the pump increase [Process, PP]

This dependencies can be semantically mapped to a domain ontology that still use concept from Brick, SSN and its extended version and Figure 2.5 shows how (part) of the discovered domain knowledge can be represented using those ontologies. It is clear that this part is the most costly of the whole process, especially timewise, since it requires very specific expertise. In the example, for instance, it is not captured that a room's internal energy is influenced by the radiant heat of the sun but it would have been possible to do so. The good side of this approach is that the modelling efforts need to be done juste once per domain as the ontology will be reusable and will adjust itself to the bulding instance trough the automatic reasoning process.

Reasoning process The reasoning process is divided in two main phases: the discovery of the physic properties and the creation of the physic processes involving those properties. This requires access to the building semantic model created during the semantic mapping phase and the physic model derived by expert for the particular domain (e.g physic model of HVAC system). Both the phases work on the ontology (metadata) to draw conclusions about the instances. The first phase has to be the the property discovery one, sicne those properties are inputs and outputs for the processes. Based on the ontology derived in the previous paragraph (??) and the semantic model of the building, through the application of simple reasoning rules, it is possible to derive the whole semantic graph of both the building and its actual physic beahviour, as seen in ??. This process is based on rules that capture domain expertise and are based on a set of entailment rules as

every room has an energy, $Room(x) \Rightarrow hasProperty(x, y) \wedge Energy(y)$

Chapter 3

Diagnosis Algorithm

Chapter 4

IoT Brain

This chapter is going to give some informations about the design choice taken while implementing the approaches in **chapter 2** and **chapter 3**. Implementation efforts were aimed at providing a further degree of scalability to the approach and leveraging cloud based technologies as well as reducing the technicality of the involved resource. It will be presented a use case that shows the

4.1 IoT BRAIN and KITT

IoT Big scale Reasoning and Analytic INsights (IoT BRAIN) allows to specify semantic models for IoT systems and run analytics across them automatically. Its core is the light-weight knowledge framework named Knowledge Inference Technology for IoT (KITT) that provides a simple API to model, manage and reason on the semantic knowledge model. It is split in multiple layers of different abstraction level that base on a common semantic modeling strategy. This architecture implements the concepts explained in the previous chapters with a strong emphasis on the scalability and availability of the approach.

IoT BRAIN, which architecture is shown in Figure 4.1, is built upon a storage layer that is comprised of a graph database, used for storing the semantic model of the building. The Watson IoT platform allows easy connection and dispatch of data securely to the cloud using the open, lightweight MQTT messaging protocol. The storage layer is completed by a dashDB, an SQL database that provides

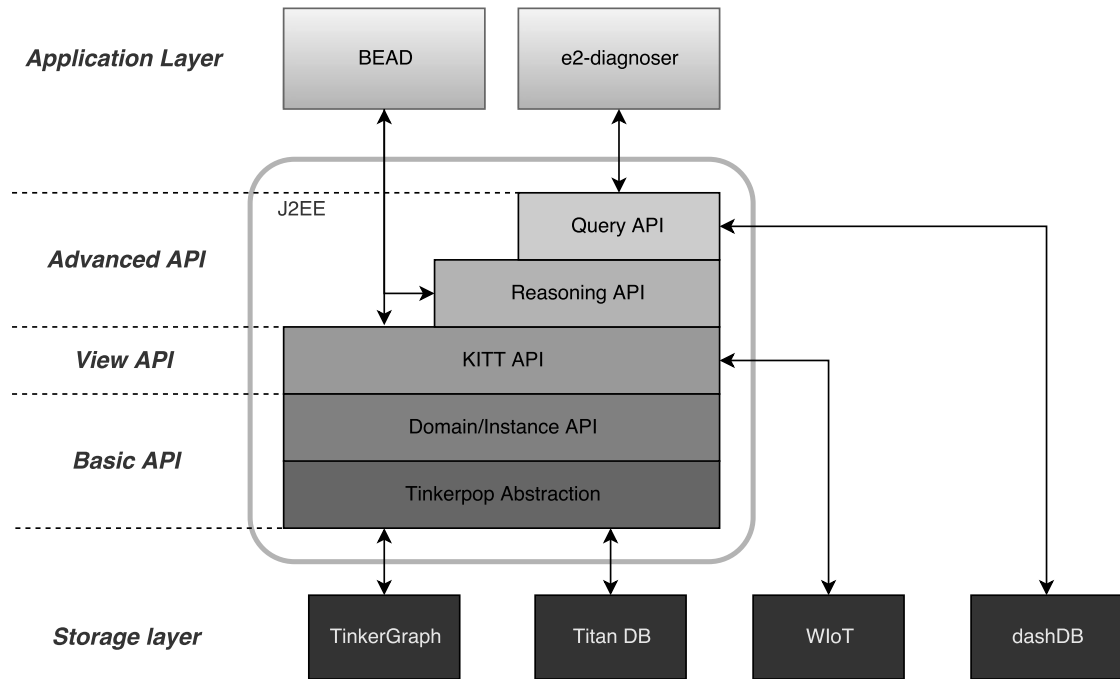


Figure 4.1: IoT BRAIN architecture

efficient storage of time series and high query performance. On top of this layer sits the core of the project which realize the theoretic process shown in Figure 2.1.

4.1.1 Graph database vs. RDF Store

why graphDB: cloud property centric need to evolve SPARQL

Bibliography

- [1] Bharathan Balaji et al. “Brick: Towards a Unified Metadata Schema For Buildings”. In: (Nov. 2016), pp. 41–50.
- [2] Michael Compton et al. “The SSN ontology of the W3C semantic sensor network incubator group”. In: *Web Semantics: Science, Services and Agents on the World Wide Web* 17 (2012), pp. 25–32. ISSN: 1570-8268.
- [3] Jingtian Cui and Shengwei Wang. “A model-based online fault detection and diagnosis strategy for centrifugal chiller systems”. In: *International Journal of Thermal Sciences* 44.10 (2005), pp. 986–999. ISSN: 1290-0729.
- [4] Srinivas Katipamula and Michael R. Brambley. “Methods for fault detection, diagnostics, and prognostics for building systems — a review”. In: *HVAC&R Research* 11.1 (2005), pp. 3–25.
- [5] Joern Ploennigs et al. “Demo Abstract: BEAD - Building Energy Asset Discovery Tool for Automating Smart Building Analytics”. In: (Nov. 2014).
- [6] *Project Haystack*. URL: <https://project-haystack.org/>.
- [7] *Resource Description Framework*. URL: <http://www.w3.org/RDF>.
- [8] Ioana Robu, Valentin Robu, and Benoit Thirion. “An introduction to the Semantic Web for health sciences librarians”. In: 94 (May 2006), pp. 198–205.
- [9] Anika Schumann, Joern Ploennigs, and Bernard Gorman. “Towards Automating the Deployment of Energy Saving Approaches in Buildings”. In: (Nov. 2014).
- [10] *SPARQL Protocol and RDF Query Language*. URL: <https://www.w3.org/TR/rdf-sparql-query>.

- [11] Yao-Wen Wang et al. “A simplified modeling of cooling coils for control and optimization of HVAC systems”. In: 45 (Nov. 2004), pp. 2915–2930.
- [12] Michael Wetter. “Multizone Building Model for Thermal Building Simulation in Modelica”. In: 2 (Jan. 2006).