Ontological Modeling of Interoperable Abnormal States

Yuki Yamagata[1], Hiroko Kou[1], Kouji Kozaki[1], Riichiro Mizoguchi[2]
Takeshi Imai[3], and Kazuhiko Ohe[3]

[1]The Institute of Scientific and Industrial Research, Osaka University
8-1 Mihogaoka, Ibaraki, Osaka, 567-0047 Japan
{yamagata,kou,kozaki,miz}@ei.sanken.osaka-u.ac.jp
[2]Research Center for Service Science School of Knowledge Science Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan
mizo@jaist.ac.jp
[3]Department of Medical Informatics, Graduate School of Medicine, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan
{ken,ohe}@hcc.h.u-tokyo.ac.jp

Abstract. Exchanging huge volumes of data is a common concern in various fields. One issue has been the difficulty of cross-domain sharing of knowledge because of its highly heterogeneous nature. We constructed an ontological model of abnormal states from the generic to domain-specific level. We propose a unified form to describe an abnormal state as a "property", and then divide it into an "attribute" and a "value" in a qualitative form. This approach promotes interoperability and flexibility of quantitative raw data, qualitative information, and generic/abstract knowledge. By developing an is-a hierarchal tree and combining causal chains of diseases, 17,000 abnormal states from 6000 diseases can be captured as generic causal relations and are reusable across 12 medical departments.

Keywords: ontology, abnormal state, property, disease, interoperability

1 Introduction

Recent advances in science and technology, together with the increasing diversity and volume of data, have led to recognition of the importance of data exchange. Scientists observe objects through experimentation to identify the attributes they possess and the values of those attributes; in other words, scientific data are represented by attributes and values. Many ontologies have been developed based on this approach, such as BFO [1], DOLCE [2], and OGMS [3], and they have contributed to understanding the semantics of heterogeneous data. However, one problem with these ontologies is the difficulty of exchanging their concepts because of their individual formulations. YAMATO has focused on integrating these differing formulations [4].

In this study, we developed an ontology of properties which allows interoperable representation in a consistent manner.

Understanding and analyzing the cause of failures and anomalies is a common issue in various domains, such as machinery, aviation, materials, and medicine.

H. Takeda et al. (Eds.): JIST 2012, LNCS 7774, pp-33-48. 2013.

Handling of appropriate knowledge representations and management of anomalies is of common concern to engineers, and knowledge sharing will contribute to the development of infrastructures for reliability/security engineering, mass-production technology, and quality assurance.

The aim of our work is to systematically develop an anomaly ontology that can capture the intrinsic properties of anomalies from content-oriented view.

In this paper, we focus on cases in the medical domain. We define abnormal states and introduce our representation model to guide the modeling of how to understand and capture the target world in the next section. Then, in Section 3, we describe our ontology of anomalies. The top level of the anomaly ontology defines very basic and generic concepts, which are "interoperable" across different domains, for example, "small in size" is interoperable between machinery and medicine. The middle-level concepts are common i.e., interoperable between diseases, e.g., "blood stenosis" in the medical domain and lower-level concepts are designed to specific-context dependent (e.g., disease-dependent) knowledge. In Section 4, we show an application in which we developed a causal chains of abnormal states covering over 6,000 diseases, capable of being used across domains by 12 medical departments, and we describe the contributions of our model to disease ontology. In Section 5, we discuss related work. Finally, we present concluding remarks and outline plans for future work.

2 Definition of Abnormal State

2.1 Basic Policy for Definition of Abnormal State

In order to understand knowledge about anomalies systematically, it is important to clearly capture essential characteristics of the abnormal states and to conceptualize them from a consistent viewpoint.

There are two different views of representations: one is content-oriented view which focuses on how to understand and capture the target world; and the other is form-oriented view, like F-logic, DL-Lite, and OWL, which focuses on dealing with how to formalize the representation of the content with syntax or axioms in logic. Our study deals with the topic of the former, i.e., content-oriented representation.

In this section, we focus on cases in the medical domain. We define abnormal states and introduce our representation model with a content-oriented viewpoint to guide the modeling of how to understand and capture ***anomalies used in definition of disease***s. A property is an abstraction of a characteristic possessed by an individual, having the value of an attribute, and is inhered in any particular entity, whereas a state is a time-indexed property that changes with time [4]. For example, when the state is “hungry” at time T1, it is represented by "being hungry" or not. Note that properties and attribute values are intrinsically different. A property such as "tall", as in "He is tall", is different from an attribute value such as "large", as in "His height is large". A property is further divided into an attribute and an attribute value.

Here, when discussing a disease in the medical domain or a failure in the engineering domain, "being in an abnormal state or not" is an important and common issue. In the medical domain, various types of representations for anomalies are used, such as "blood glucose level 250 mg/dL, "glucose concentration is high", and "hyperglycemia".

Based on YAMATO [4], which is an upper ontology that was carefully designed to cover both quality and quantity ontologies, we classified abnormal states into the three categories shown in Table 1.

Table 1. Representation of anomaly

Representation	Example	Usage
quantitative representation	blood glucose level 250mg/dl blood pressure 200 mmHg	Diagnostics
qualitative representation	glucose concentration is high blood pressure is high	
property representation	hyperglycemia hypertension	Definition of disease

In many diagnostic situations, a quantitative representation is indispensable, because identifying a precise value for each patient by clinical tests is the main task. However, the definition of a disease is different. In the above three instances, "being hyperglycemic" might be a common and intrinsic nature. Actually, as explained in many textbooks or guidelines, to illustrate disease, most abnormal states are described as property representations; for example, to explain diabetes, "Diabetes is characterized by hyperglycemia caused by impaired insulin action..." [5].

In this way, to capture the essentials of abnormal states, a property representation seems to be the most natural and intrinsic. Therefore, first, we captured abnormal states as properties, represented by a tuple like <Property (P), Property Value (Vp)>. Basically, Property Value takes a Boolean value, i.e., <existence / non-existence>. For example, "constriction (stenosis)" is described as <constriction, existence>. In addition, when necessary, a degree value[1] can also be used for describing the degree of the Property Value.

An ontology is a theory of concepts which explain the target world. When explaining abnormal states, it says "having abnormal values". Abnormal states also can be explained by some kind of disturbance of the homeostasis in human body. While the former does not necessarily have connotation of bad for health, the latter does. Therefore, making a decision about the latter is not the job of ontology researchers but rather is the one of medical experts based on the judgment of the medical knowledge. For example, answering a question whether the high HDL cholesterol[2] level is "abnormal state" or not in the latter sense is not a task of ontologists but that of medical experts, which is not discussed in this article.

[1] The degree value of Property Value is, for example, "mild/moderate/severe". For example, <diarrhea (P), severe (Vp)>.

[2] HDL cholesterol is known as good cholesterol in medicine.

In practice, however, a property is too abstract to represent the precise meaning of data. In the case of the above "constriction", it would have a more concrete meaning, say, <cross-sectional area, small>.

Therefore, we expanded the specification of a property by decomposing it into a tuple: <Attribute (A), Attribute Value (V)>. We adopt Attribute Value in a qualitative form, i.e., Qualitative Value (Vql). This approach contributes to promoting consistency in representation, as well as interoperability between quantitative raw data and generic/abstract anomaly knowledge (see Section 2.3).

In the engineering domain, most properties can be decomposed into a set of <A, V>, whereas in clinical medicine, some properties cannot be decomposed, because the precise mechanisms in the human body remain elusive. In such cases, for example, deformation and pain, the property representation could be an undecomposed one: <Property (P), Property Value (Vp)>. Whether such abnormal states represented in terms of properties defined above makes sense or not is dependent on the advance of medicine.

One of the advantages of representation in terms of property is that it is substantially unaffected by small parameter fluctuations. Another advantage is that representation in terms of property in an abstract form makes it easier to capture the essentials of each disease. Although a diagnostic task requires a quantitative representation rather than a qualitative representation in terms of property, this is clearly a different task from defining disease.

2.2 Anomaly Representation

2.2.1 Standard Representation

In the human body, abnormal states are highly diverse. They have a variety of grain sizes, from cell and organs to action level. In addition, related agents involved with these states, such as chemicals, viruses, bacteria, and so on, are also extremely diverse. In this section, we apply our representation model to medical abnormal states and examine whether we can represent them appropriately and consistently.

Because an attribute cannot exist alone but always exists in association with an independent object, it is necessary to identify what it inheres in (hereinafter referred to as “target object”). For example, in the case of "intestinal stenosis", the target object of its attribute “cross-sectional area” is neither the blood vessels nor other organs but the intestine (Fig. 1, upper left).

In the case of "hypertension", blood is the target object of “pressure”. In order to represent the target object of an attribute, we introduce "Object" and decompose a property into a triple: <Object (O), Attribute (A), Attribute Value (V)>. For example, "gastric dilatation" and "hypertension" are decomposed into <stomach, volume, large> and <blood, pressure, high>, respectively. This is a standard representation model of anomalies.

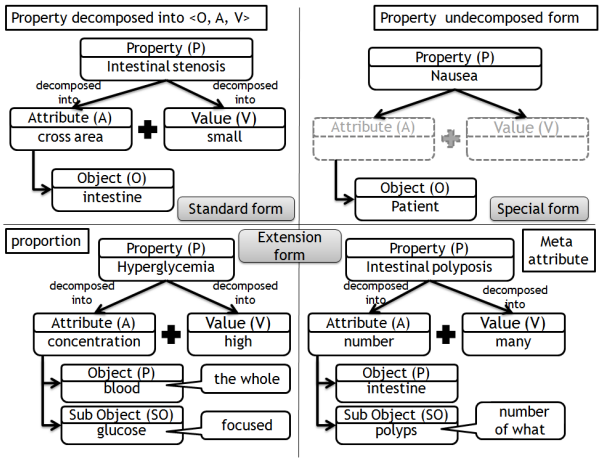

Fig. 1. Framework for abnormal states in medical domain

2.2.2 Advanced Representation

Although many properties are represented in terms of the standard triple, some are not. These are properties defined by a ratio and a meta-attribute and need a more advanced representation.

A) Ratio

In the case of a representation involving a ratio, the target object having the ratio must be identified. In addition, an advanced representation for what will be focused on ("focused object") is needed. For example, in the case of a representation for "hyperglycemia", concentration is a known attribute of blood, so blood is identified as Object. However, since concentration means the ratio of the focused object relative to the whole mixture, glucose is needed for representation as the focused object. Therefore, we introduce Sub-Object (SO) to represent a focused object.

There are different kinds of ratio, depending on what is focused on. Therefore, Object and Sub-Object vary according to the kind of ratio as follows:

m/n ratio (no unit): Represented by the ratio of the focused object (m) relative to the whole (mixture) (n), which has no unit:

In this kind of ratio, the whole mixture (n) is identified as Object having a ratio as its attribute, and focused object (m) is defined as Sub-Object, which represents "the ratio of what". For instance, since blood glucose concentration is the amount of glucose (sugar) present in the blood, it is a property of blood as the whole mixture, and the focused object is glucose. Accordingly, the representation of "hyperglycemia" is a quadruple, <blood (O), glucose (SO), concentration (A), high (V)> (Fig. 1, bottom left), where blood is defined as Object, and glucose is Sub-Object (Table 2, No. 1).

Table 2. Representations used with ratio

No.	Proportion Classification	Property (P)	Value	Attribute (A)	value (V)	Object (O)	Sub Object (SO)	Ratio
1	m/n (focused on m/ the whole)	high m ratio	existence	ratio	high/ low	the whole	focused	m/n
	example	hyperglycemia	existence	Glucose concentration	high	blood	Glucose	Glucose/ Blood
2-1	*m/n* (focused on m)	high *m* ratio	existence	ratio	high/ low	object	m	m/n
	example	high Albumin ratio	existence	Albumin ratio	high	urine	Albumin	Albumin/ Creatinine
2-2	*m/n* (focused on the ratio)	high/low *m/n* ratio	existence	ratio	high/ low	object	blank	m/n
	example	increased A/G ratio	existence	A/G ratio	high	blood	blank	Albumin/ Globulin

2- m/n ratio (of same object): Represented by the ratio of attribute m to attribute n of the same object:

2-1 m/n (focused on m): Represented by the ratio of the focused attribute (m) to another attribute (n) of the same object:

The thing having both attributes m and n is identified as Object, and the focused attribute m is defined as Sub-Object. For example, Urine Albumin-to-Creatinine Ratio (UACR) is the ratio of albumin level to creatinine level in the urine. The albumin is Sub-Object, which is focused on, and urine is defined as Object. Therefore, "high UACR" is represented as a quadruple <Urine (O), Albumin (SO), Albumin ratio (A), high (V)> (Table 2, No. 2-1).

2-2 m/n ratio (focused on the ratio): Defined by the ratio of attribute m to attribute n of the same object:

In this kind of ratio, what is focused on is simply the ratio itself rather than either attribute. For instance, in the Albumin to Globulin Ratio (A/G ratio), which is the ratio of the albumin level to globulin level in the blood, blood is defined as Object, though Sub-Object is null since what is focused on is neither the albumin nor globulin level but the A/G ratio itself. Thus, "decreased A/G ratio" is described as a triple: <blood (O), A/G ratio (A), low (V)> (Table 2, No. 2-2). Accordingly, in this kind of ratio, the thing having both attributes m and n is identified as Object, whereas Sub-Object is null.

B) Meta Attribute

In the property "intestinal polyposis", color and size are attributes of polyps. However, "many polyps" is not an attribute of "polyps" because it is not an inherent nature.

Following the meta-attribute in YAMATO, in which “number of curves” is identified as a meta-attribute of a road which has many curves, we regard "the number of polyps" as a meta-attribute of the intestine. By introducing "Sub-object", the property "intestinal polyposis" is decomposed into a quadruple <intestine (O), polyps (SO), number (A), many (V)>, where intestine is identified as Object, and polyps are described as Sub-object, which collectively represent "number of polyps" (Fig. 1, bottom right).

Applying the concept of Sub-Object provides a flexible representation in the case of complicated concepts like ratio and meta-attribute and allows us to determine Objects as distinct targets with consistency. As shown in Fig. 1, various descriptions for abnormal states are supported by our framework.

In summary, the key point of our model is, first, to capture an abnormal state as a property and introduce a property representation framework, and second, to decompose the representation into a standard triple <Object (O), Attribute (A), Attribute Value (V)>. This framework ensures a consistent representation of various simple abnormal states. Furthermore, introducing the concept of "Sub-Object" is effective in increasing the flexibility and consistency of representations in complicated cases, such as those involving ratios and meta-attributes, described as a quadruple <Object (O), Sub-Object (SO), Attribute (A), Attribute Value (V)>.

2.3 Interoperability between Property and Attribute

In medical institutions or hospitals, huge volumes of diagnostic/clinical test data have been accumulated and stored, most of which are quantitative data: e.g., blood pressure (systolic) 180 mm Hg. By using our representation model, quantitative data can be described as <Object (O), Attribute (A), Quantitative Value (Vqt)>. For example, in the case of "blood pressure (systolic) 180 mm Hg", it is decomposed into <blood (O), pressure (A), 180 mmHg (Vqt)>.

Here, a threshold[3] based on a generic value used in the domain is introduced, and a qualitative value can be obtained by comparing the threshold value with quantitative value. For example, above the threshold (e.g. 130mmHg), the Quantitative Value (180mmHg) is converted to the qualitative value "high".

Therefore, the quantitative representation <blood (O), pressure (A), 180 mmHg (Vqt)> can be converted into a qualitative representation <blood (O), pressure (A), high (Qualitative value (Vql))>, which is a decomposition of the property representation for "hypertension". Accordingly, to deal with raw data, a quantitative representation as <Object (O), Attribute (A), Quantitative Value (Vqt)> can be used.

In Section 2.2, we introduced the property representation framework <Object (O), Property (P), Property Value (Vp)>, which is decomposed into its qualitative representation <Object (O), Attribute (A), Qualitative Value (Vql)>. This enables interchangeability from a quantitative representation to a property representation. For example, <blood (O), pressure (A), 180 mmHg (Vqt)> can be exchanged with <blood

[3] How to set the threshold and identifying the precise value is a diagnostic task at the instance level. We do not discuss them further in this paper.

(O), hypertension (P), severe (Vp)>, which is sufficient to be judged as "being an anomaly" or not. Therefore, our approach contributes to promote interoperability between quantitative raw data, such as clinical examination data, and generic/abstract anomaly knowledge.

That is to say, our anomaly representation model is classified into three interoperable groups as follows:

1.1. Raw data representation :<O, A, Vqt>
1.2. Abnormal state representation: <O, P, Vp>
1.3. Specific abnormal representation: <O, P, Vp, So>.

3 *Is-a* hierarchy of Anomaly Ontology

In this section, we present our approach for building a framework for knowledge systematization using ontological engineering.

In general, domain experts work with strongly domain-specific knowledge, consequently making it more difficult to share common and generic knowledge. To cope with this problem, in developing our anomaly ontology design, we have to cover both specific concepts (i.e., context-dependent concepts) and basic/generic concepts across multiple domains, namely, context-independent concepts.

To build a framework for knowledge systematization using ontological engineering, we propose the following three-layer ontological model of abnormal states with an *is-a* hierarchical tree (Fig. 2):

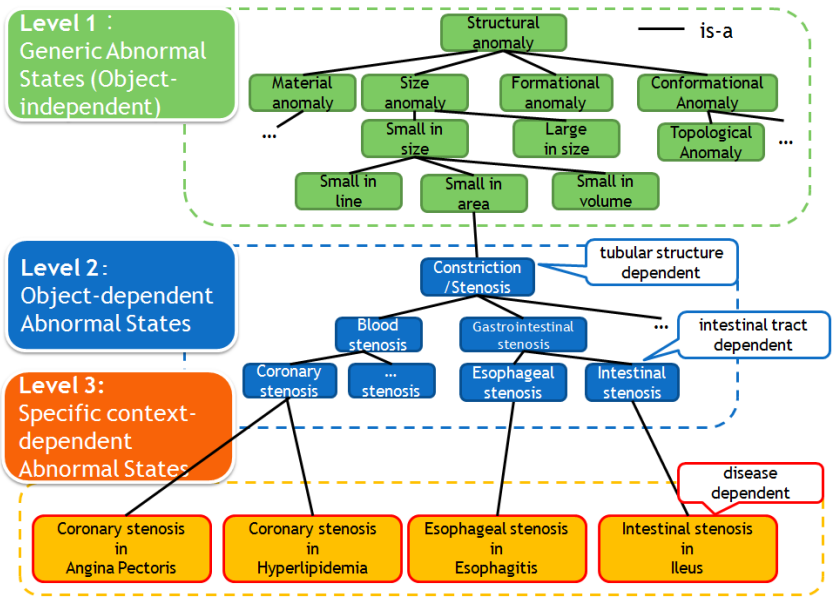

Fig. 2. Three-layer ontological model of abnormal states

1- Top layer (Level 1): Generic abnormal states
2- Middle layer (Level 2): Object-dependent abnormal states
3- Bottom layer (Level 3): Specific context-dependent abnormal states

The details are discussed in the following subsections.

3.1 Level 1: Generic Abnormal States

The top level of the anomaly ontology defines and provides very basic and generic concepts commonly found in several objects, such as cracking, deformation, color changes, dysfunction, and so on. They do not depend on any structures; in other words, they are primitive states[4] that are very general and usable in a wide range of domains, such as machinery, materials, aviation, and medicine.

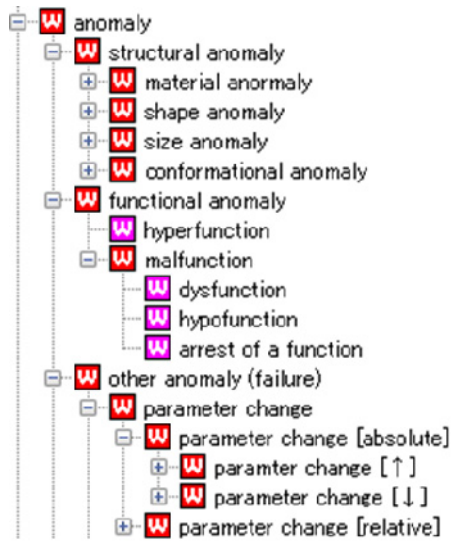

Fig. 3. Top-level categories related to an abnormal state

Consider any abnormal states in multiple domains. Generally, they can be classified into two main groups: structure-related and function-related. Therefore, in this study, the top-level category of generic abnormal states has three subclasses: "structural anomaly", "functional anomaly", and "other anomaly" (Fig. 3). A structural anomaly is defined as an abnormal state associated with structure. It is sub-categorized into material anomaly (e.g., hardening), formation anomaly (e.g., deformation), size anomaly, and conformational anomaly, such as topological anomaly (e.g., dislocation), structural defects (e.g., adduct/loss of parts) etc., while still retaining the identity of the structural body.

Generic abnormal states are context-independent states found in several objects, i.e., object-independent. They are represented in a recursive structure, with reference to Level 2 abnormal states. For example, "small in area" is defined as a lower-level size anomaly at Level 1, which is referred to in various domains and is specified as Object-dependent concepts at Level 2.

A functional anomaly is defined as an abnormal state related to failure to function properly. It is classified into hyperfunction and malfunction, the latter of which is subcategorized into dysfunction, hypofunction, and function arrest, respectively.

[4] In this paper, we deal with a single primitive state. A composite state, composed of more than one single state, e.g., "torsional vibration", which is composed of torsion and vibration, is omitted due to space limitation.

Other abnormal states include parameter anomalies, which are classified into increased or decreased parameter, depending on whether or not the attribute has a higher or lower value than a threshold level. Various generic abnormal states related to parameter change, e.g., increased/decreased velocity, increased/decreased pressure, etc., are defined by using general physical parameters, such as velocity, pressure, and concentration.

3.2 Level 2: Object-Dependent Abnormal States

Top level concepts at Level 2 are dependent on generic structures, such as "tubular structure", "bursiform structure", etc., which are common and are used in many domains. By identifying the target object and specializing generic abnormal states of Level 1, an *is-a* hierarchal tree at Level 2 was developed with consistency. For instance, by specializing "small in area" at Level 1, "constriction of tubular structure", where the cross-sectional area has become narrowed, is defined at Level 2, and this is further specialized in the definitions "water pipe narrowing", "intestinal stenosis", and so on.

In the lower level of the tree, abnormal states dependent on domain-specific objects are defined. In the medical domain, for example, "constriction of tubular structure" at the upper-left in Fig. 4 is specified as "vascular stenosis" dependent on "blood vessels" (Fig. 4, middle left.), and is further specified into "coronary stenosis" dependent on coronary artery (Fig. 4, lower left.). Thus, we can represent concepts at all required granularities in the expert domain.

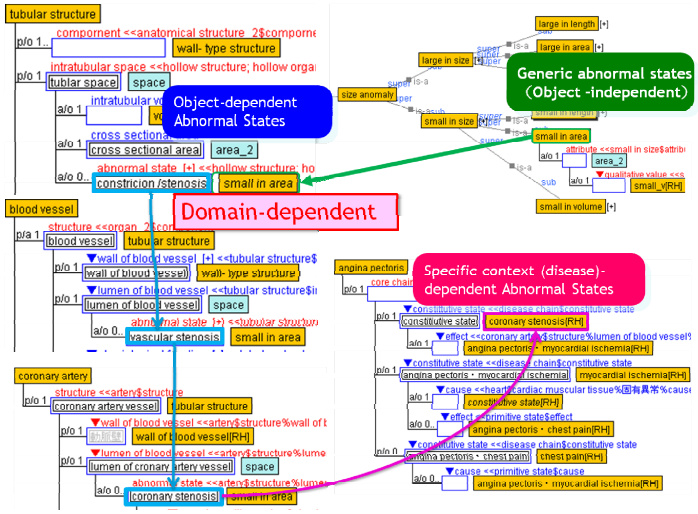

Fig. 4. Specialization of abnormal states from generic level to specific level

Here, in general, the difficulty arises in how fine a granularity needs to be supported in our ontology. In the medical domain, for example, as compared to "polyp formation", "gastric polyp formation" in a specific organ (the stomach) might be redundant and dispensable. However, it is widely known that abnormal states on one component subsequent influence adjacent components, which causes other abnormal states. Actually, a medical concept like gastric polyp is different from nasal polyp; the former

causes gastric cancer, whereas the latter causes dysosmia. Therefore, there is a need for distinct abnormal states at specific organ levels.

From an ontological engineering point of view, our framework for modeling abnormal states is intended to capture abnormal states from the generic level to the specific domain-level, so as to provide abnormal states in a necessary number of specific organ/tissue/cell layers in the medical domain.

Note that, although such abnormal states of specific cells are defined at Level 2, they are distinct from context-dependent concepts at Level 3.

For example, hypertension, which means high blood pressure, can be defined in a context-independent manner at Level 2 with reference to level 3 concepts used in various diseases. For instance, how hypertension causes renal artery sclerosis should be defined in a context-dependent manner, i.e., a disease-dependent manner (e.g., renal artery sclerosis-dependent) at Level 3.

3.3 Level 3: Specific Context-dependent Abnormal States

Level 3 concepts are captured as context-dependent abnormal states, which are specialized from abnormal states at Level 2 in specific diseases or machine failure. For example, "coronary stenosis" dependent on coronary artery at Level 2 is defined as a constituent of the disease angina pectoris at Level 3, which causes myocardial ischemia (Fig. 2, lower right), which is also defined as a constituent of hyperlipidemia.

In summary, our ontology can represent various anomalies with consistency. In our ontological approach, common concepts can be kept distinct from specific ones and can be defined as appropriate according to their context. By building an anomaly ontology with an *is-a* hierarchy tree, higher level concepts are more generic and can be shared as cross-domain common knowledge, and lower-level concepts are designed to represent domain-specific knowledge for the required granularity, which is suitable for practical expertise. Consequently, our approach can provide backbone concepts with a machine understandable description, which supports the development of an infrastructure for anomaly knowledge of failures/diseases and will be useful for a wide range of applications.

4 Application of Anomaly Ontology

We have been developing a disease ontology in our Japanese Medical Ontology project. We define a disease as a dependent continuant constituted of one or more causal chains of clinical disorders appearing in a human body and initiated by at least one disorder [6].

Each disease is structured as a directed acyclic graph (DAG), which is composed of nodes (clinical disorders) and relations. We can introduce an *is-a* relation between diseases using chain-inclusion relationship between causal chains as follow:

Disease A is a super type of disease B if the core causal chain of disease A is included in that of disease B. The inclusion of nodes is judged by taking an *is-a* relation between the nodes into account, as well as sameness of the nodes. For example, that diabetes and type-1 diabetes are respectively defined as <*deficiency of insulin* → *elevated level of glucose in the blood*> and <destruction of pancreatic beta cells → lack of insulin in the blood → deficiency of insulin → elevated level of glucose in the blood>. Then, we get <*type-1 diabetes is-a diabetes*>.

Currently, clinicians from 12 medical departments have described causal chains consisting of over 6000 diseases and approximately 17,000 abnormal states (Table 3).

We defined upper level concepts (classes) such as clinical disorders (abnormal states), causal chains, causal relationships (cause and effect), etc. based on YAMATO. In addition, by using these abnormal states, we developed an anomaly ontology with a three-layer model, as described in the previous section, and have also applied our representation model as described in Section 2.

Table 3. Statistics

Medical Department	Number of Abnormal state	Number of Disease
Allergy and Rheumatoid	806	101
Cardiovascular Medicine	2,289	550
Diabetes and Metabolic Diseases	1,989	445
Orthopedic Surgery	1,121	208
Respiratory Medicine	1,739	788
Neurology	1,893	397
Ophthalmology	1,306	561
Nephrology and Endocrinology	868	196
Hematology and Oncology	354	415
Dermatology	908	1,086
Pediatrics	2,334	879
Otorhinolaryngology	1,118	470
Total	16,725	6,096

Abnormal states are defined as causes or results of each disease, i.e., context-dependent concepts at Level 3 described in one medical department. For instance, in the Department of Gastroenterology, one clinician defines "intestinal stenosis" as the cause of ileus. Each Level 3 abnormal state is stored as an organ-dependent abnormal state in the Level 2 leaf. Furthermore, at upper levels, abnormal states are more sharable. For example, "intestinal stenosis" and "esophagostenosis" are able to share "stenosis of digestive tract" at Level 2. Moreover, "coronary stenosis" can be shared with a corresponding upper level 2 concept, i.e. "tubular stenosis", via the *is-a* hierarchical tree of the anomaly ontology.

Level 2 leaf abnormal states are able to be reused as reference information for other diseases at all 12 medical departments. For example, level 3 "coronary stenosis" described in the Department of Cardiovascular Medicine is first stored as a Level 2 leaf abnormal state, and later it can be reused as a result of Level 3 "accumulation of cholesterol" in hyperlipidemia at another department, namely, the Department of Diabetes and Metabolic Diseases (Fig. 5).

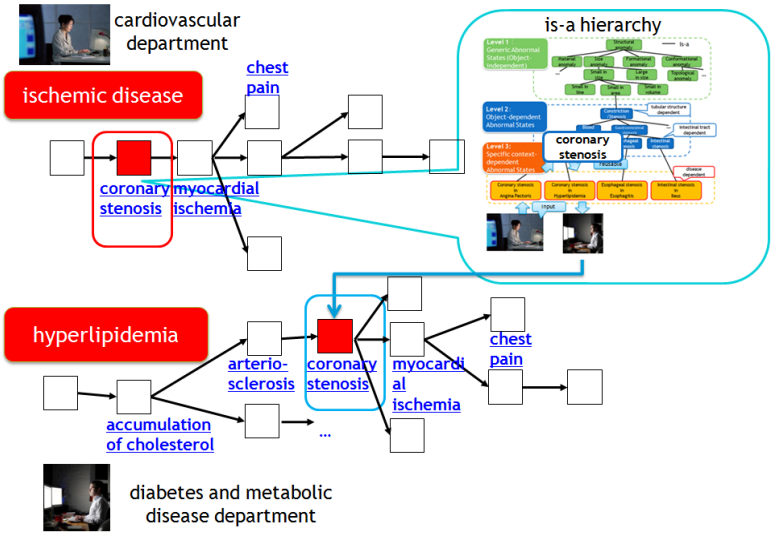

Fig. 5. An example of sharing of abnormal states between different medical departments

Note that, once a causal chain related to an abnormal state is described, another clinician can use it as a generic causal chain. Consequently, a clinician at the Department of Diabetes and Metabolic Diseases can annotate the generic progression of "coronary stenosis".

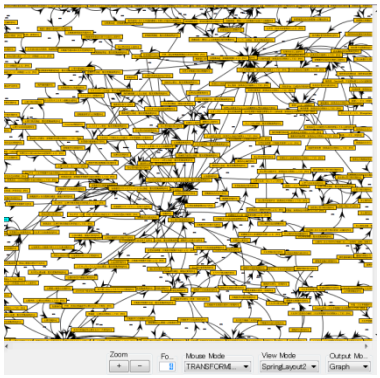

Fig. 6. An example of the dynamic generation of generic causal chain of a disease

Though the causal chains of each disease are described at particular medical departments, generic causal chains can be generated by combining causal chains that include the same abnormal states, which allows all causal relationships, including 17,000 abnormal states from 12 medical departments, to be visualized (Fig. 6).

Since our approach can manage both the causal relations of diseases and the related *is-a* hierarchical structure of the anomaly ontology, it might be a good infrastructure for managing various kinds of anomaly knowledge.

In this sense, using our infrastructure, which connects abnormal states to various granularities of abnormal states via (1) causal chains of disease i.e., horizontal relations among the same level, and (2) an *is-a* hierarchy in the ontological structure, i.e., vertical relations among the different level, will facilitate cross-domain usage among heterogeneous concepts of abnormal states and diseases/failures. This method should aid in understanding the mechanisms that cause diseases and lead to more-complete understanding of related abnormal states. Furthermore, knowledge regarding canonical causal chains of upper-level abnormal states can be annotated

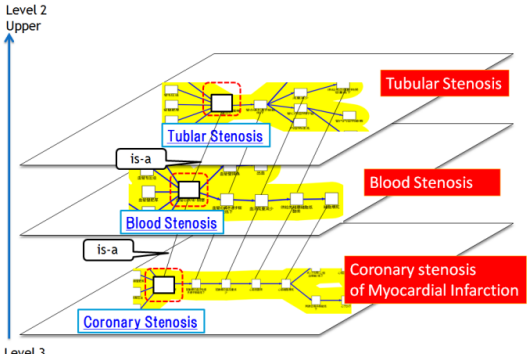

Fig. 7. Causal chain layers

from the relations among real causal chains by using the anomaly ontology and causal chains (Fig. 7).

For example, from a causal chain including "coronary stenosis" at Level 2, we could obtain an upper-level causal chain including "blood stenosis" and the even higher level "tube stenosis".

If we trace the vertical relation of each abnormal state from some causal chains of diseases that are understood, a common scenario might be generated as a canonical causal chain, which will provide a clue to revealing the mechanisms of other diseases that remain poorly understood.

5 Related Work

There are at least three methods of describing qualities in BFO [1], DOLCE [2] and Galen [7]. BFO recommends the <Entity, Property> (e.g., <John, tall>) formalism, whereas DOLCE recommends <Entity, Attribute, Value> (e.g., <John, height, 180cm>) and Galen recommends <Entity, Property, Value> (e.g., <John, tallness, large>). However, all three descriptions have the same meaning, namely, John's height.

A property description such as <bone, deformed> seems not to be covered by DOLCE whereas it is covered by BFO, which does not cover <Entity, Attribute, Value>. The YAMATO ontology covers all these three kinds of descriptions and supports interoperability among them.

Since it is based on the YAMATO ontology, our representation model is not only able to manage the three kinds of description but can also be used as a reference ontology for representing properties (qualities) among these upper ontologies.

In the biotechnology community, Phenotypic Quality (PATO) [8] is a famous ontology of phenotypic qualities, defining composite phenotypes and phenotype annotations, and the current version of PATO adopts property descriptions, e.g., <eye, red>. Our ontology is compatible with the latest PATO description, and also with older versions of EAV (Entity + Attribute + Value) .

In the medical domain, many medical ontologies have been developed for realizing sophisticated medical information systems, such as Ontology for General Medical Science (OGMS) [3], DO [9], IDO [10]. However, they do not include sufficient information regarding the relations between abnormal states in one disease. In future, we plan to link to an external ontology such as OGMS and provide useful information about causal relationships in diseases so that the concepts complement each other.

6 Conclusions

In this paper, first, we discussed various complicated issues of abnormal state concepts and introduced a representation model that handles 17,000 abnormal states from approximately 6000 diseases by capturing properties, which are consistently decomposed into attributes. Next, we proposed a systematic method for developing an anomaly ontology from the generic level to the specific domain level. We have also developed a hierarchical tree for capturing both commonality and specificity discretely. Our model demonstrated that common knowledge about abnormal states can be shared at the upper level of abnormal states and is reusable for describing other diseases in different medical departments.

From preliminary studies on abnormal states from our ontological tree, the number of top level (Level 1) generic abnormal states, which are not dependent on the human body and are common concepts in multiple domains, is about 100.

Furthermore, with a combination of our anomaly ontology and causal chains of diseases, we can capture all causal relations of 17,000 abnormal states in approximately 6,000 diseases from 12 medical departments.

In the medical domain, e-health needs data exchange, such as Electronic health records (EHR), in order to exchange data appropriately, and it is necessary not only to manage data as quantitative values but also to capture the intrinsic essentials as semantics. Hence, our representation model has interoperability between qualitative data and the properties of anomalies, as shown in Section 2.3, which might contribute to organizing an integrated system in which various anomalies from raw data to abstract knowledge are accessible and manageable by computers. Our approach provides various useful information for a better understanding of the essentials of abnormal states in disease, and in addition, provides practical usability from the raw data level to the semantics level.

As shown in Fig. 8, from the raw data level to the anomaly knowledge level, a three-level architecture that consists of (1) a database for raw clinical data, (2) a

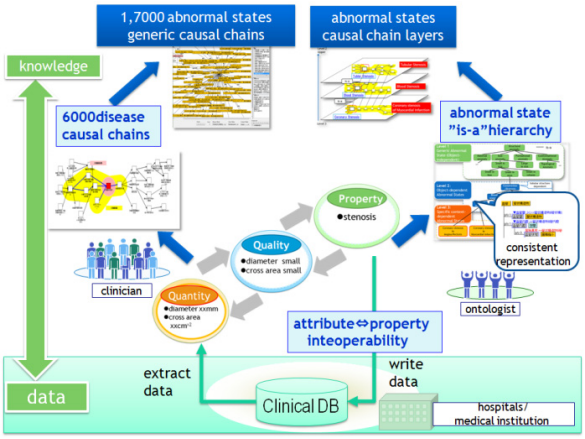

Fig. 8. Integrated system architecture for anomaly knowledge

processor of interoperating quantitative, qualitative and property information, and (3) a knowledge space for abnormal states, where an ontology and causal chains are developed as a backbone, might be suitable.

Acknowledgements. A part of this research was supported by the Ministry of Health, Labor and Welfare, Japan, through its "Research and development of medical knowledge base databases for medical information system" and by the Japan Society for the Promotion of Science (JSPS) through its "Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)".

References

1. Grenon, P., et al.: Biodynamic Ontology: Applying BFO in the Biomedical Domain. In: Pisanelli, D.M. (ed.) Ontologies in Medicine, pp. 20–38. IOS, Amsterdam (2004)
2. Guarino, N.: Some Ontological Principles for Designing Upper Level Lexical Resources. In: Proc. of International Conference on Lexical Resources and Evaluation (1998)
3. Scheuermann, R.H., Ceusters, W., Smith, B.: Toward an Ontological Treatment of Disease and Diagnosis. In: Proc. of the 2009 AMIA Summit on Translational Bioinformatics, San Francisco, pp. 116–120 (2009)
4. Mizoguchi, R.: YAMATO: Yet Another More Advanced Top-level Ontology. In: Proceedings of the Sixth Australasian Ontology Workshop, pp. 1–16 (2010)
5. Kuzuya, T., et al.: Report of the Committee on the Classification and Diagnostic Criteria of Diabetes Mellitus. Journal of the Japan Diabetes Society 53(6), 450–467 (2010)
6. Mizoguchi, R., et al.: River Flow Model of Diseases. In: Proc. of ICBO 2011, pp. 63–70 (2011)
7. OpenGALEN, `http://www.opengalen.org/`
8. Gkoutos, G.V., et al.: Ontologies for the description of mouse phenotypes. Comp. Funct. Genomics 5, 545–551 (2004)
9. Osborne, J.D., et al.: Annotating the human genome with Disease Ontology. BMC Genomics 10(1), S6 (2009)
10. Cowell, L.G., Smith, B.: Infectious Disease Ontology. In: Sintchenko, V. (ed.) Infectious Disease Informatics, ch. 19, pp. 373–395 (2010)