A Method for Making Contextual Features Explicit in Domain Ontologies

Mariela Rico[1], Ma. Laura Caliusco[1], Ma. Laura Taverna[1], Omar Chiotti[1,2], and Ma. Rosa Galli[1,2]

[1] CIDISI Research Center, UTN-FRSF, Lavaise 610, S3004EWB Santa Fe, Argentina
{mrico,mlcaliusco}@santafe-conicet.gov.ar,ltaverna@fiq.unl.edu.ar
[2] INGAR-UTN-CONICET, Avellaneda 3657, S3002GJC Santa Fe, Argentina
{chiotti,mrgalli}@santafe-conicet.gov.ar

Abstract. Ontologies are used to represent the semantics of entities of reality in several application domains, leading to different domain ontologies. In order to enable semantic integration of heterogeneous information sources using ontologies, appropriate matching techniques are required. In the last years, different techniques have been proposed. However, these techniques work poorly when there are incomplete representations of entities, which is usually the case in practice. With the aim of improving the representation of entities in an ontology, this paper proposes a method for making features, whose interpretation depends on the considered context, explicit. In addition, an application example is discussed.

1 Introduction

With the widespread use of ontologies, different parties are inevitably developing domain ontologies with overlapping contents. Different methods and methodologies for building ontologies have been tailored to suit the requirements of various application domains [1]. They provide a means of identifying the relevant entities and relationships in the domain of interest. The representation of these entities is highly reliant on the intended uses of the ontology. Hence, the way in which an entity in reality is represented may differ depending on the set of requirements the ontology should satisfy. When ontologies developed in this way are later matched with other ontologies the alignment between them is often error-prone.

Different matching techniques to help deal with the ontology matching problem have been proposed [2]. These techniques have proven to be very effective when the input ontologies are rich in details, but they work poorly when the available representations are incomplete, which is often the case in practice.

With the aim of providing a better basis for matching, recent work has been proposed. In [3], a methodology for enriching class hierarchies with ontological information is presented, which supports approximate matchings of class hierarchies plainly based on subsumption reasoning. In [4], it is shown that improving and systematizing the naming of concepts can improve matching. Since these proposals do not provide a thorough representation of entities, this paper focuses on making these representations more precise by making the contextual features of entities explicit.

Previous work has shown that making the contextual features of an entity explicit significantly improves the results of the ontology matching process [5]. The objective of this paper is to show the advance made in this area, which consists of a method for making these features explicit. To this aim, the paper is organized as follows. Section 2 describes the proposed method. Section 3 presents an application of the method. Finally, Section 4 is devoted to the conclusions.

2 A Method for Making Contextual Features Explicit

The aim of the proposed method is to improve the representation of entities, whose instances must be matched, by making their contextual features explicit in the ontology. The processes that compose the method are described next.

2.1 Process 1: Identify the Entities Represented in the Ontology and their Features

Any improvements on the representation of entities involve prior knowledge about the entities to which that representation relates. The objective of this process is to clarify what the entities, their features, and their relationships with other entities are, according to the way they are represented in the ontology.

This first process allows the ontologist to identify the features, which are generally implicit in the representation of an entity, and although they may be inferred by a human agent, they cannot be inferred by a machine agent. These features are not required to be made explicit when the entities are considered within the same context, but it becomes necessary when the entities have to be interpreted in another one. Here, a context is viewed as a reference environment in which descriptions of real world objects are given [6].

The outputs of this process are both a list of the entities and their relationships, and, for each entity, a list of their features, whose semantics is affected by the context in which the entity is considered. The list of the entities can be obtained exploring the documentation associated with the ontology. The list of the features can be made by encoding the knowledge required to understand the corresponding entities. For example, street, street number, floor, apartment, city, postal code, province or state, and country are features of a mailing address and that differentiate it from an e-mail address. If necessary, domain experts could help to identify the features by indicating the underlying knowledge they assume should be possessed in order to have an accurate interpretation of the meaning of the entities in different contexts.

2.2 Process 2: Identify the Ontology Elements that Represent the Entities

In this process, terms, relations between terms, axioms, etc., used to represent each entity and their features must be identified as recognized in Process 1. Depending on the perspective, an entity could have been represented with only a simple term or with a set of representational elements.

2.3 Process 3: Identify the Features that Have to be Made Explicit

When representing an entity, it is possible that some of its features are implicit or the representation of these features is incomplete. Therefore, their explicitness becomes necessary for improving the representation of the real entity semantics. Based on the outputs of processes 1 and 2, these features can be detected. Since not all of these features need to be made explicit, answering the following questions could help to identify the ones that do need.

- Are there any implicit features in the representation of an entity that although they may be inferred by a human agent they cannot be inferred by a machine agent? If the answer is yes, could these features be inferred in the wrong way in other contexts different from the considered context? If the answer is yes, these features should be made explicit.
- Are there any entities whose representations and/or meanings could change depending on the context in which the entities are considered? If the answer is yes, are the representations and meanings of the features or entities completely explicit in the ontology? If the answer is no, the representations and meanings should be made explicit.
- What are the quality dimensions used to represent a feature? Are they the same regardless of the context in which the feature is considered? If the answer is no, are they explicit in the ontology? If the answer is no, these quality dimensions should be made explicit.

2.4 Process 4: Make the Features of Each Entity Explicit

Once the features and their representation dimensions have been identified, they have to be made explicit. For this purpose, reusing existing and widely accepted ontologies should be considered. This can be made by importing the ontology to be reused as will be shown in Section 3.4. Examples of such ontologies are:

- The OWL-Time ontology for expressing facts about topological relations among instants, intervals, and events, together with information about durations and about dates and times [7].
- The ISO 3166 Country Codes Ontology for modeling the official country names (http://www.daml.org/2001/09/countries/iso-3166-ont).
- The UN/CEFACT ontology for modeling physical quantities and units of measurement (http://www.unece.org/cefact/codesfortrade/codes_index.htm).

When it is not possible to reuse an ontology, it is necessary to identify if the feature is simple or complex. A simple feature is a quality that does not bear other qualities, and it is associated with a one-dimensional representation in human cognition [8]. For example, the weight of a thing is associated with a one-dimensional structure, whose possible values are positive real numbers. Thus, two elements should be added to the ontology: a term denoting the representation dimension, and a relation between this term and the term that represents the simple feature.

A complex feature is a quality that bears other qualities, and it is associated with a set of integral dimensions that are separable from all other dimensions [8]. An integral dimension is one in which it is not possible to assign a value to an object on one dimension without giving it a value on the other. For instance, color can be represented in terms of the dimensions of hue, saturation, and brightness. These dimensions are integral. By contrast, weight and hue dimensions are said to be separable. Each integral dimension is associated with a simple feature. In order to improve the representation of a complex feature, the following elements should be added to the ontology:

- A term representing the set of integral dimensions and a relation between this term and the term that represents the complex feature.
- For each integral dimension, a term representing it and a relation between this term and the term that represents the set of integral dimensions.
- For each term representing an integral dimension, a relation between this term and the term that represents the corresponding simple feature.

In addition, for each term representing a one-dimensional representation or an integral dimension, a term representing the unit of measurement of the dimension and a relation between these two terms should be added to the ontology. This term affects the granularity of the dimension but not its structure. For example, a weight dimension has positive real numbers as values regardless of whether the metric unit is kilogram or ton.

2.5 Process 5: Designate a Bridge Term that Refers to Each Entity

The intended uses of an entity in the context considered should be represented by terms, called bridge because they allow linking different meanings and representations of the same entity in different contexts. These terms should also be interpreted as representing contextual features, because the intended use depends on the context in which the entity is considered.

Thus, in the ontology, it is necessary to determine if there is a term that designates the intended use of each entity in the considered context, but if such term is absent, it has to be added. These bridge terms should also be related to the elements that represent the entity whose intended use they represent. An entity could be represented by a single element or a set of elements. In the former case, a relation between that single element and the bridge term should be added. In the latter case, the most representative term should be chosen and then, a relation between this term and the bridge term should be added. As the bridge term represents a contextual feature, it should also be related to the term that represents its representation dimension.

3 An Application Example

Suppose there is a collaborative relationship between a packaging industry (supplier) and a dairy industry (customer). Both trading partners have to interchange the information shown in Table 1 to reach an agreement on a replenishment plan.

Table 1. Example of the necessary information to decide on a replenishment plan.

Horizon: 6/04 - 31/05					
Period	Product				Quantity
	Id	Trademark	Type	Size	
7/04 - 13/04	20320101	yy	Carton	1000	4400
	20320102			2900	2880
	20070231		Plastic	196	1600
	20070232			250	1800
	20320101	zz	Carton	1000	2200
	20320102			2900	8064
	20070232		Plastic	250	1800
	20070235			1000	6500
14/04 - 20/04	...	...	...	...	...

The structure and semantics of this information is initially reflected in an ontology called *EBD ontology*, which is shown in Fig. 1. In order to improve the representation of entities in this ontology, the proposed method is applied.

3.1 Process 1: Identify the Entities and their Features

The entities, whose information has to be translated, are: (i) Trading partners that assume two different roles: supplier and customer. A relevant feature for a trading partner is its address. (ii) The replenishment plan that refers to the agreed plan between the trading partners. Some of the features of a replenishment plan are: the time period during which the plan is valid, the products which are exchanged, the quantities of them, and the different periods within the horizon during which these products are exchanged, among others. (iii) The products involved in the replenishment plan are the ones manufactured by the packaging industry, and the packages for the dairy industry products. Some of the features of a product are its trademark, type, and size.

3.2 Process 2: Identify the Ontology Elements

A trading partner assuming the supplier role is represented by the terms `Agent`, `Organization`, and `Supplier`, plus their properties, and the `isa` relations between these terms. A customer is represented by means of the terms `Agent` and `Organization`, their properties, and the relation `isa` between them. Their addresses are represented by the term `Address` and the relation `hasAddress`.

The entity replenishment plan and its features are represented by the terms `EBD`, `EBDItemsCollection`, and `EBDItem`, their properties, and the relations `has Items` and `hasItem`. The term `EBD` refers to the documents that are interchanged between the trading partners. The terms `EBDItemsCollection` and `EBDItem` represent the structure of the documents.

The products and their features are represented by the following properties of the term `EBDItem`: `PartNumber`, `ItemName`, and `ItemDescription`.

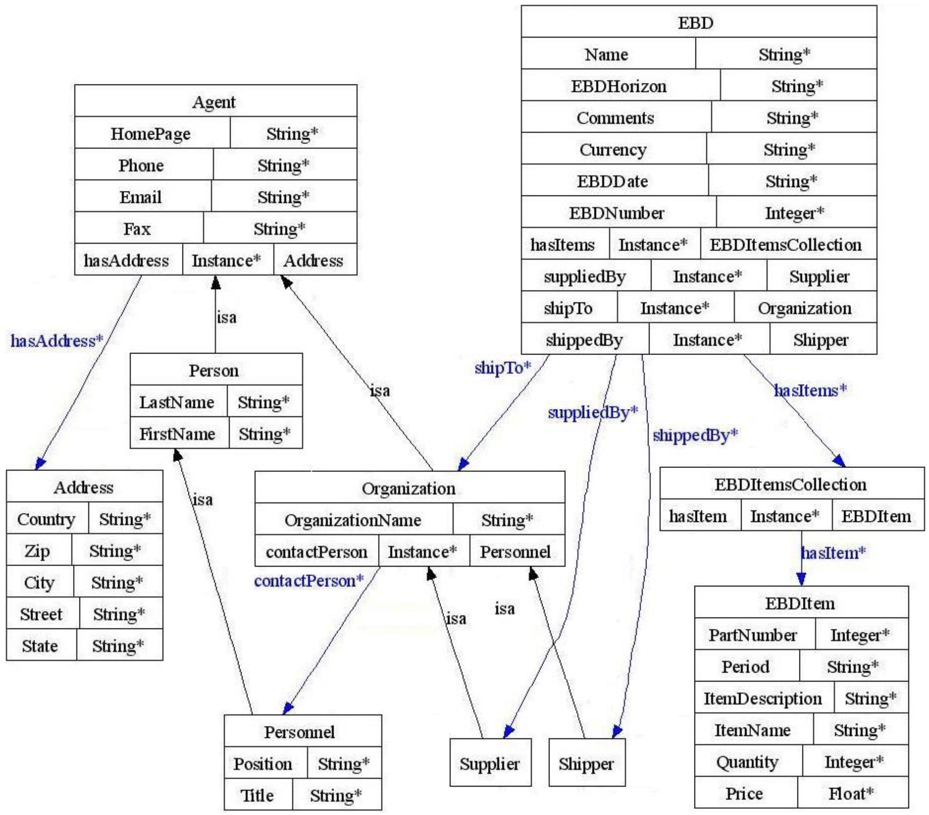

Fig. 1. The original EBD ontology.

3.3 Process 3: Identify the Features that have to be Made Explicit

As mentioned above, a relevant feature for a trading partner is its address, which is represented by the term `Address`. According to Smith [9], this term represents a quality existing in reality. This quality refers to a mailing address and not to an e-mail address, for example. The features of a mailing address are street, street number, floor, apartment, city, postal code, province or state, and country. In the ontology, not all of them are explicit such as floor and apartment. However, they do not prevent a correct interpretation of the entity in any context in which that entity is considered. A reason for making these features explicit could be to have a more complete representation of the entity.

Considering the replenishment plan, the time period during which the plan is valid is represented by the property `EBDHorizon`, whose data type is `string`. This way of representation does not reveal if horizon is expressed as an amount of time or as an interval, thus not satisfying the minimal encoding bias criterion [10]. According to Table 1, the representation should make the feature explicit as a calendar interval since the horizon is not only a quantity of time, but it also has a location on the time line. A similar analysis can be made about the periods within the horizon (represented by the property `Period` of the term `EBDItem`).

Another feature of a replenishment plan is the ordered quantity of each product. This feature is represented by the property `Quantity`, whose data type is `integer`. At this point, some questions arise. In what unit of measurement is this quantity expressed? This could refer to units of products or units of product packs, for example. Now, will any information system that has to deal with such information make a correct interpretation of it? If there is any possibility of misunderstanding, the unit of measurement should be made explicit.

The products are represented by the following properties of the term `EBDItem`: `PartNumber`, whose data type is integer; and `ItemName` and `ItemDescription`, whose data type is string. `PartNumber` represents the Product Id of Table 1, but `ItemName` and `ItemDescription` not represent, at first sight, any of the other product features. `ItemDescription` can refer to a description of the product in natural language. However, the property `ItemName` should be replaced by three terms that represent the features: trademark, type, and size.

3.4 Process 4: Make the Features of Each Entity Explicit

In order to make a feature explicit, it is important to distinguish when that feature is an entity in itself and when it is not. For example, the feature country of the address could be considered an entity too, and be represented by means of the term `Country` instead of a property. In this case, a formal relation [8] that glues together the terms `Address` and `Country` is needed. By contrast, the feature floor is existentially dependent on the address in the context under consideration. Then, a property is more appropriate to represent it.

Taking the OWL-Time ontology [7] into account, the feature horizon of the entity replenishment plan should be represented by means of a different term derived from the `CalendarClockInterval` term, and linked to the `EBD` term by a formal relation. The same treatment applies to the periods within the horizon.

Since the ordered quantity of each product is a simple feature, four elements should be added to the ontology: (i) the term `QuantityDimension` denoting the representation dimension, (ii) the term `UnitOfMeasure` representing the unit of measurement of the dimension, (iii) a relation between these two terms, and (iv) a relation between the term `QuantityDimension` and the term that represents the simple feature. However, quantity is not represented by a term but by a property. Although it is possible to think of the quantity not as an entity but as a property, it is necessary to represent it by a term to follow the minimal encoding bias criterion [10]. Additionally, this term, called `Quantity`, has to be related to the term `EBDItem` by a formal relation. Fig. 2 shows a portion of the ontology after making the quantity and temporal features explicit.

In order to represent the features of products, three terms are added: `Type`, `Size`, and `Trademark`. The term `Trademark` is related to another term, `Trademark Dimension`, which represents the representation dimension of the feature trademark. This representation dimension is an enumeration of possible values. In the same way, the term `Type`, which represents the type of material from which

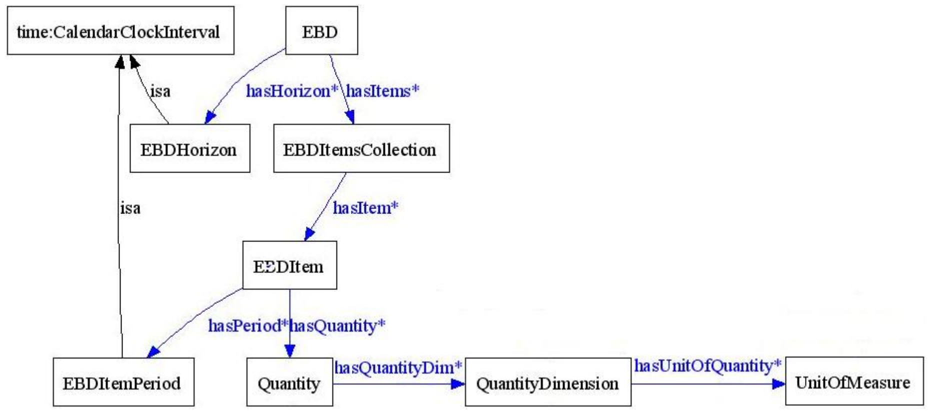

Fig. 2. A portion of the extended EBD ontology representing temporal and quantity features.

the packages are made, has a relation with the term `TypeDimension`. This term represents the type dimension, which is also an enumeration of possible values (such as "carton" and "plastic"). Finally, the term `Size`, which represents the capacity of the packages, is related to the term `SizeDimension`, which represents the representation dimension of the feature size. This dimension is metric, i.e. its possible values are in the set of non-negative numbers (196, 250, etc.). Since the capacity of the packages is an amount associated with a unit of measure, the term `SizeDimension` has a relation with the aforementioned term `UnitOfMeasure`.

So far, only the features of the entity product have been represented. However, the entity product lacks a representation of its own. To this aim, the term `Product` has to be added to the ontology. This term has to be related to the terms `EBDItem`, `Trademark`, `Type`, and `Size`. Additionally, the properties `PartNumber` and `ItemDescription` have to be properties of the term `Product` but not of the term `EBDItem`.

Besides being an entity in itself, a product is a feature of the entity replenishment plan. As such, it is a complex feature. Thus, the term `Product` is related to a set of representation dimensions represented by the term `ProductMultiDimension`. This set is called multi-dimension and has the property of having a value assigned to each dimension. The dimensions that compose this multi-dimension are called integral dimensions [8]. In the case of `ProductMultiDimension`, such dimensions are represented by the terms `TrademarkDimension`, `TypeDimension`, and `SizeDimension` aforementioned.

3.5 Process 5: Designate a Bridge Term that Refers to Each Entity

An important feature that should be made explicit is the intended use of an entity in the considered context. However, trading partners are not regarded as physical entities but as roles, which they assume. The treatment of entities that represent roles is postponed to future work.

The intended use of the entity replenishment plan is to represent the agreed plan by the trading partners. Since the term `EBD` refers to the documents that are interchanged between the trading partners, particularly the replenishment plan, it can be used to designate the intended use of the entity.

The products involved in the replenishment plan can be misunderstood depending on their intended use. They are the ones manufactured by the packaging industry, and the packages of the dairy industry products. Then, in an ontology from the collaborative relationship context such as the EBD ontology, two bridge terms should be added: `Product` and `Package`, representing the intended use of the entity in the packaging industry context and in the dairy industry context respectively. Since `Product` is already in the ontology, only `Package` has to be added and related to the term `Product`. As bridge terms represent contextual features, they have to be also related to the terms that represent their representation dimension in human cognition. In the dairy industry context, packages are associated with a representation dimension, which is an enumeration of possible values. Fig. 3 shows a portion of the EBD ontology with the changes made to represent the feature product adequately.

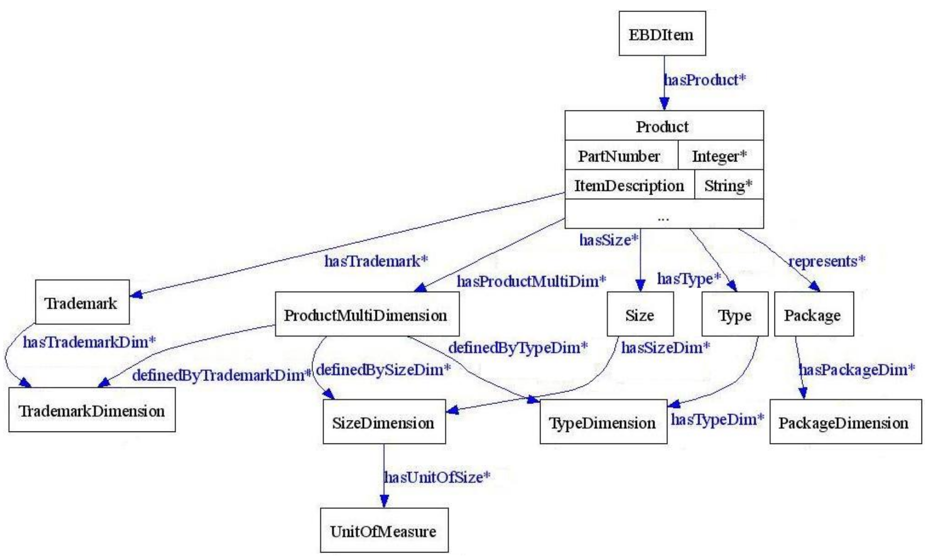

Fig. 3. A portion of the extended EBD ontology representing products.

4 Conclusions and Future Work

In this paper, a method for making the representation of entities in an ontology more precise was presented. To this aim, making the contextual features of entities explicit was proposed. This explicitness improves the results of the ontology matching process as has been shown in previous work [5].

In practice, ontologies suffer from different kinds of modeling errors independently of the context in which they are used. The proposed method can also be applied to existing ontologies to overcome some of these errors.

With the purpose of supporting the explicitness of the features, a metamodel based on the modeling principles of Model Driven Architecture has been proposed [11]. At present, a prototype, which implements such metamodel, is being developed to support the method proposed in this paper.

Acknowledgments. The authors are grateful to "Universidad Tecnológica Nacional (UTN)", "Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)", and "Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT)" for their financial support.

References

1. Cristani, M., Cuel, R.: Methodologies for the semantic web: state-of-the-art of ontology methodology. Column of SIGSEMIS Bulletin. Theme: SW Challenges for KM. **01**(02) (2004) 103–112
2. Euzenat, J., Shvaiko, P.: Ontology matching. Springer-Verlag, London (2007)
3. Stuckenschmidt, H., van Harmelen, F.: Information Sharing on the Semantic Web. Advanced Information Processing. Springer Verlag, Berlin (2004)
4. Sváb-Zamazal, O., Svátek, V., Meilicke, C., Stuckenschmidt, H.: Testing the impact of pattern-based ontology refactoring on ontology matching results. In Shvaiko, P., Euzenat, J., Giunchiglia, F., Stuckenschmidt, H., eds.: Proc. of the 3rd Int. Workshop on Ontology Matching (OM-2008). (2008)
5. Rico, M., Caliusco, M.L., Chiotti, O., Galli, M.R.: Toward the improvement of alignment between heterogeneous domain ontologies. In: Proc. of the 34 CLEI, Santa Fe, Argentina. (2008)
6. Theodorakis, M.: Contextualization: An Abstraction Mechanism for Information Modeling. PhD thesis, Department of Computer Science, University of Crete, Crete, Greece (2001)
7. Hobbs, J.R., Pan, F.: An ontology of time for the semantic web. ACM Transactions on Asian Language Information Processing (TALIP) **3** (2004) 66–85
8. Guizzardi, G.: Ontological foundations for structural conceptual models. PhD thesis, University of Twente, Enschede, The Netherlands (2005)
9. Smith, B., Kusnierczyk, W., Schober, D., Ceusters, W.: Towards a reference terminology for ontology research and development in the biomedical domain. In: Proc. of the 2nd Int. Workshop on Formal Biomedical Knowledge Representation: "Biomedical Ontology in Action". (2006) 57–66
10. Grüber, T.: Toward principles for the design of ontologies used for knowledge sharing. Int. J. Hum.-Comput. Stud. **43**(5-6) (1995) 907–928
11. Rico, M., Caliusco, M.L., Taverna, M.L., Chiotti, O., Galli, M.R.: Collaborative commerce in supply chains: An approach to support semantic interoperability. In: Proc. of the 37 JAIIO-JII, Santa Fe, Argentina. (2008)