A model for formalizing characteristics in Protégé-OWL

Anna Estellés y Amparo Alcina[1]

[1] Tecnolettra Team, Universidad Jaume I,
{estelles, alcina}@trad.uji.es

Abstract: This paper proposes a model for formalizing concept characteristics in a manner consistent with the Terminology theoretical framework. We will focus on the identification of the elements that take part in characteristic formalization and their role, as well as on the double nature of characteristic. Finally, we provide a preliminary outline of characteristics description formalization using an ontology editor called Protégé-OWL. This article shows the preliminary work of a broader project that deals with the representation of characteristics and its terminological information using Protégé-OWL. It is addressed to terminology researchers and developers that may use ontologies in order to represent specific domain conceptual systems.

Keywords: Characteristic, concept, formalization, Terminology, Ontology.

1 Introduction

In Terminology, concepts can be classified in four categories or groups: entities, activities, relations and characteristics (Sager y Kageura, 1994:192). Our goal is to describe characteristics by means of a formal language used to build ontologies (OWL-DL) and an editing tool called Protégé-OWL. The context of this work are Txtceram and ONTODIC[1] projects, so we base our research in examples of tile characteristics. In order to describe a characteristic it is necessary to explore its nature, the categories into which it can be grouped, the domain of expertise in which the characteristic is used, its relevance for various concepts descriptions, and how it relates to other concepts.

The use of ontologies in applied terminology has become an alternative to traditional terminological databases. This has lead to the emergence of various studies within the discipline of Terminology (Gruber, 1993; Gamper, Nejdl *et al.*, 1999; Temmerman & Koen, 2003). When using ontologies as a resource for representing concept

1 «TXTCeram: Semi-automatic extraction and analysis of ceramics terms using Electronic corpora» and «ONTODIC: Methodology and tecnologies for creationg onomasiological dictionaries ontology based». Project director: Amparo Alcina. http://tecnolettra.uji.es

systems from a terminological point of view, for instance, for creating an onomasiological dictionary, ontology seeks the representation of all the concepts in a domain and not only entities (Alcina, 2008:34).

Next section focuses on the elements that contribute to the description of a characteristic from a terminological point of view. Finally, we outline the possibilities of formalizing these elements using Protégé-OWL.

2 Characteristic definition

In the normative document UNE EN 1066 (UNE:1066, 1991:3) characteristic are defined as: «*Characteristics serve as a basis for classifying concepts. They are necessary for differentiating one concept from another in a specific area and also for other functions*». A similar definition could be: «*an element of a concept denoting a property of an entity or a class of entities*» (Madsen, 1998:342).

This definition answers the question 'What is a characteristic for?' but it does not provide satisfactory answers to the question 'What is a characteristic?'. Nuoponnen's definition of characteristic answers this question stating that a characteristic is «*a concept, the referent of which is a property of an entity*» (Nuopponen, 1994:61).

Above definitions show that on the one hand, for a concept system, a characteristic is a property of a concept, one of the most important features for the description of a concept. On the other, characteristics are an independent concept.

In the field of ontology and knowledge engineering (Masolo & Borgo, 2005), a *quality* is an entity that is inherent in another entity. The weight of a book is an example. The value of a quality is named a *quale*, the 'actual' weight of a particular shaft is the quale. A quale is considered an abstract entity that is part of a quality region, where other quales compose the quality space. An example of color quality space would be all the quales of colour (blue, red...).

Meyer (Meyer *et al.*, 1997:100) and Kageura (Kageura, 2002:76), among others, explore this approach.

Based on this and other bibliography of Terminology, we distinguished between the different elements that make up a characteristic description, and their roles in the description. First, we delimited the two components of a characteristic: category (characteristic label) and value. Then, we took into account that a characteristic can be grouped using classes. This led us to the domain of a characteristic, the concept or concepts that can be described by a characteristic; thereafter we focused on the subject, the context in which a characteristic is used in discourse.

Characteristic label: is the category that serves to group and identify the values which may be attributed to the characteristic. It is also used as an abstract concept when it is not followed in the discourse by a value, where the label is supposed to be understood as the concept representing all the possible values of a characteristic. An example of characteristic label is 'colour'.

Characteristic value: is a specific entity that designates an aspect of a concept. We distinguish among the ones that are expressed lexically and the rest. RED or CHINESE RED are examples of lexical characteristic values. In our analysis, we also found the following types of non-lexical values: numbers (integers, float or percentage) + measurement unit; symbols (such as GHA, GHB, GHC - values of 'chemical resistance' in a tile) and Boolean (for instance, YES or NO for 'frost resistance').

Classes and subclasses of characteristic labels and values will be grouped according to their commonalities. These commonalities are independent of the concept they describe, for example, 'size' and 'height' are characteristics used when measuring things, and they could be part of a class called 'measurement'. An example of classification of values is shown in Kageura's work (Kageura, 2002:77).

Domain: An important relation exists between a characteristic and the concept or concepts it describes. We already mentioned the example of the characteristic 'frost resistance' which has a relationship with 'tile'. *Domain* is the concept or group of concepts described by a characteristic. Notice that a characteristic can apply to more than one domain, even in the same context and that if a characteristic applies to a domain, concepts within this domain can also be described by that characteristic.

Area: refers to the area of specialization of a characteristic. When building a terminological database, terminographers normally enter a field that refers to the specialized area of the terms they work on. This provides explicit information, and enables future merging.

3 Outline of formalization and representation of a characteristic[2] in Protégé-OWL

Protégé OWL is an editing tool based on OWL (*ontology web language*). In OWL there are different elements: `owl:individual` is used to represent domain objects, `owl:class` represents a group of individuals or a concept and `rdf:property` is used for expressing types of relationships between instances or classes. The `rdf:property` element is linked to: a domain, `rdfs:domain` (concept or concepts described by the property) and a *range,* `rdfs:range` (value or values of the property). Fuhermore, there is the `owl:ObjectProperty` element that links an individual or class to an individual or class. The `owl:DatatypeProperty` element links an individual or class to an XML Schema Datatype value (numer, boolean value, symbol). Finally, using `owl:annotationProperty` we can add information to classes, individuals or properties though this information can not be processed by a reasoner (Sanz & Jiménez-Ruiz, 2009).

3.1 A characteristic in Protégé-OWL

For our project, first we have represented the characteristic label by means of OWL class or an individual. These classes or individuals are to be part of the taxonomy of characteristic labels or categories.

The area is formalized using an *annotation property*. Furthermore, we have formalized a characteristic as a descriptive element by using OWL *properties.* Once they have been formalized, *properties* are located in individuals or class descriptions.

If characteristic values are represented in a lexical form, we represented them first making use of classes and subclasses in the characteristic taxonomy. Then, we linked the value or values to the object of description by using an *object property* (fig. 1).

If characteristic values are not a lexical form, we used *datatype properties*. For example, 'glaze' is a characteristic with Boolean values, then the datatype property ‘has_glaze’ has a domain 'tile' with a Boolean range, Protégé default Boolean values are TRUE or FALSE.

Notice that in OWL there can not be more than one class or property with the same name, when needed we recommend to use prefixes and suffixes such as ‘colour_value’, ‘has_colour’ and ‘colour_prop_of'.

2 In order to avoid reader’s confusion on the different approaches to the concept ‘property’, we will use characteristic when referring to a 'property' from a terminological point of view (Nuopponen, 1994:61), property when referring to OWL properties (Horridge, 2009:23) and feature when referring to a functional element.

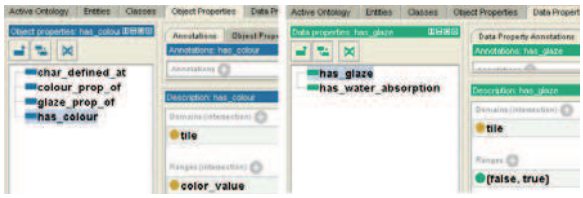

Fig. 1 - *Example of characteristic formalization as an element for describing concepts*

To describe that a 'characteristic label' represents the same concept as a *property*, and describes a specific concept or concepts, we used a new *object property.* This is done using OWL syntax. For example, 'glaze_prop_of' is a property whose domain is 'glaze' and range is "has_glaze *value* TRUE" (fig.3). By applying a reasoner such as Fact++ we infer that a range of this property whose domain is the characteristic label 'glaze' (class) is the entity concept 'glazed_tile' (class).

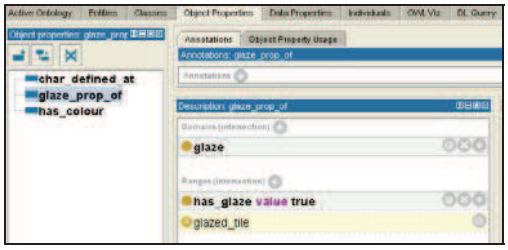

Fig. 2 - *Example of property establishing a link between the characteristic class and the concept described*

4 Conclusions and future work

In this paper, we have set boundaries on the elements involved in the description of a characteristic and their roles. We are currently working on the formalization of tile characteristics following this description and testing the representation of characteristic labels by means of instances. Since some lacks have been detected regarding the 'translation' of characteristic in Terminology using knowledge representation concepts, future work is motivated by a further research in this area. Furthermore, characteristics have a number of properties related to their role in concepts description, for example characteristic can be essential, complementary (Cabré, 1992:184) or distinctive (ISO:704, 2000:5), as well as extrinsic or intrinsic (Cabré, 1992:185). Formalization of those properties and other semantic relations forms part of a future work we are developing nowadays. Further results will be presented in the form of a PhD Thesis.

References

ALCINA, A. (2009). «Metodología y técnicas para la elaboración de diccionarios onomasiológicos». In Alcina, A., E. Valero & E. Rambla (Eds.), *Terminología y Sociedad del conocimiento*. Berna, Peter Lang.

CABRÉ, M. T. (2000). «Terminologie et linguistique: la théorie des portes.» *Terminologies Nouvelles*(21): 10-15.

GAMPER, J., W. NEJDL & M. WOLPERS (1999). *Combining ontologies and terminologies in information systems*. Proceedings of the 5th International Congress on Terminology and Knowledge Engineering, Innsbruck, Austria, TermNet.

GRUBER, T. (1993). «A translation approach to portable ontology specification.» *Knowledge Acquisition* **5**(2): 199-220.

HORRIDGE, M. (2009). A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools. (Edition 1.2). C.-O. project, The University of Manchester.

ISO:704 (2000). ISO 704: Terminology work -- Principles and methods.

KAGEURA, K. (2002). *The Dynamics of Terminology: a descriptive theory of term formation and terminological growth*. Amsterdam, John Benjamins.

MADSEN, B. N. (1998). *Typed featured structures for terminology work - Part I*. LSP - Identity and Interface - Research, Knowledge and Society. Proceedings of the 11th European Symposium on Language for Special Purposes, Copenhagen.

MASOLO, C. & S. BORGO (2005). Qualities in formal ontology. In *Workshop on Foundational Aspects of Ontologies (FOnt 2005)*, Koblenz, Germany.

MEYER, I., E., KAREN & D. SKUCE (1997). «Systematic Concept Analysis within a Knowledge-Based Approach to Terminology». *Handbook of Terminology Management*. S. E. B. Wright, Gerhard. Amsterdam, John Benjamins. **1:** 98-118.

NUOPPONEN, A. (1994). «On causality and concept relationships». *Terminology Science and Terminology Planning, IITF-Workshop on Theoretical Issues of Terminology Science*. J. K. Draskau y H. Picht Vienna, TermNet**:** 217-230.

SAGER, J. C. & K. KAGEURA (1994). «Concept Classes and Conceptual Structures: Their Role and Necessity in Terminology.» *ALFA* **7/8**: 191-216.

SANZ, I. & E. JIMÉNEZ-RUIZ (2009). «Ontologías en Informática». *Terminología y sociedad del conocimiento*. A. Alcina, E. Valero y E. Rambla. Berna, Peter Lang**:** 255-286.

TEMMERMAN, R. & K., KOEN (2003). *Termontography: Ontology Building and the Sociocognitive Approach to Terminology Description*. CIL17, Praga, Matfyzpress, MFF UK (CD-ROM).

UNE:1066 (1991). UNE 1066: Principios y métodos de la terminología. AENOR.