

Integrating Two Descriptions of Taxonomies with Materialization

Alain Pirotte, Université catholique de Louvain, Institut d'Administration et de Gestion, Louvain-la-Neuve, Belgium.

David Massart, European Schoolnet Office, Brussels, Belgium.

This paper presents a precise correspondence between two views of taxonomic hierarchies: an intensional view based on concepts and an extensional view based on categories, i.e., subsets of the population of individuals analyzed in terms of these concepts. The correspondence is described with materialization, a generic relationship defined for object-oriented and entity-relationship information models. The paper introduces materialization and shows how it provides a systematic bridge between both views of taxonomies.

1 REAL-WORLD MODELING

We view the world of interest as consisting of *objects* (i.e., things, individuals, ideas, ...) important enough to be distinguished from one another. Still, for clarity, we will say that the world is populated by *individuals* and we will reserve the term *object* for the usual basic construct of object-oriented models.

The world can be described in many ways. We are interested in precise intensional descriptions, called *schemas* in the database culture, based on *concepts* that are ideas or notions of various degrees of generality about individuals and sets of individuals in the world. Concepts are used, for example, for distinguishing individuals from other individuals or for characterizing the common properties of similar individuals. *Categories* are the extensional counterparts of concepts. They serve to classify the population of individuals in the world into subsets perceived as interesting.

The activity of *conceptual modeling* builds such intensional descriptions, also called conceptual models. They are biased and incomplete symbolic images of portions of the world built with concepts. Conceptual models capture meaning in a processable form, in order to perform various tasks of symbolic manipulation regarded as useful (e.g., understand, aggregate, transform information; generalize available information; make assumptions and explore their consequences).

Taxonomies are conceptual models. Informally, a *taxonomy* is an organization of concepts or of categories about individuals in the world structured by an order relation expressing the relative generality of the concepts or categories.

2 TWO DESCRIPTIONS OF TAXONOMIES

Taxonomies can be described as hierarchies structured along two alternative dimensions:

- **a dimension of concepts**, that structures the population of individuals in the world in terms of concepts organized along an intensional dimension of abstractness/concreteness. That is the method usually chosen to organize scientific knowledge (e.g., biological organisms).

The hierarchy of concepts may be expressed as a hierarchy of classes and their instances (or metaclasses and their classes) structured by the mechanism of classification of usual object-oriented models. Concepts are progressively refined by successive instantiations downwards in the hierarchy;

- **a dimension of populations**, that characterizes a collection of subsets of individuals in the world as a hierarchy of classes and subclasses structured by the generalization abstraction of usual object-oriented models. The dimension of populations analyzes the overall population of interesting individuals in terms of smaller and smaller subsets downwards in the hierarchy.

Consider, for example, the population of vehicles on the road in Belgium.

Figure 1 shows a view of the example based on concepts¹. This is the point of view, for example, of the transportation board or of the accounting office. These agencies are not interested in individual vehicles, but rather in the structure of tax revenues, or in the regulation for driving licences and car insurance. In that view, the class of the most general concepts is **Types of vehicle**. It has three instances: **class A vehicle** (or truck), **class B vehicle** (or car), and **class C vehicle** (or bus), each concept being characterized, for example, by a value for a type of driving license and for a type of insurance. The concept of **car** is in turn refined into concepts of **luxury car**, **family car**, and **sports car**, which are instances of class **Types of car** associated with the **class B** concept. Then, the concept of **family car** is in turn refined as car models, which are instances of class **Types of family car**.

Figure 2 shows the same example in terms of categories of world individuals, namely sets of concrete vehicles². This is the point of view, for example, of the registry office that issues individual driving licences or vehicle plates, and collects taxes. In that view, the top-level class **Vehicles** comprises all the vehicles of interest. It has three subclasses (**Trucks**, **Cars**, and **Buses**), denoting the corresponding subsets of vehicles, and so on. Figure 2 also shows two subclasses of class **Family cars**, distinguished by their model: **Fiat Retro cars** and **2CV cars**, and two instances of these classes, **Guy's 2CV** and **Nico's fiat**, denoting real concrete cars.

¹Classes are drawn as rectangular boxes and objects as rounded boxes; class names begin with an uppercase letter, whereas object names are written in lower case. Dashed lines denote instantiation links (classification).

²Solid lines denote generalization links.

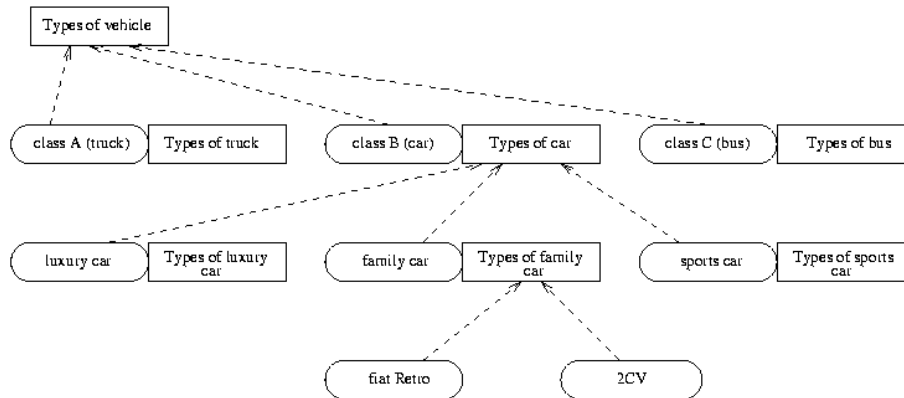


Figure 1: Concept dimension of a taxonomy of vehicles

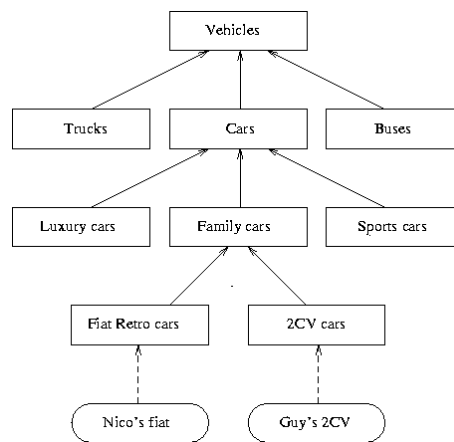


Figure 2: Population dimension of a taxonomy of vehicles

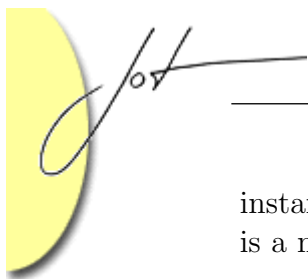
3 TWO-FACETED CONSTRUCTS

In the population view of the taxonomy, the properties of each class are derived from its links with its superclasses through the inheritance mechanism of generalization.

In the concept view, each concept *c* (like *class B* or *car*) is an object, which is tightly bound to a class (*Types of car*) whose instances (such as *luxury car*) are objects denoting subconcepts of *c*. The taxonomy of concepts is thus expressed as a hierarchy of classes structured by the classification link of object-oriented models. Concepts are viewed alternatively as objects and as classes of their subconcepts.

Two-faceted constructs make that double explicit. Each two-faceted construct is a composite structure associating an object and a class. The association is underlined by drawing each two-faceted construct as a class box adjacent to an object box.

For example, in Figure 1, objects *luxury car*, *family car*, and *sports car* are instances of class *Types of car*, each object being associated with a class in a two-faceted construct (e.g., concept *family car* is associated with class *Types of family car* in a two-faceted construct). Similarly, *fiat Retro* and *2CV*, which denote car models, are



instances of Types of family car. In object-oriented terms, Types of car, for example, is a metaclass for objects fiat Retro and 2CV.

Thus, each concept is a two-faceted construct with an object facet (a concept is an instance of a more abstract concept at the next higher level of the taxonomy) and a class facet (a concept is a class of refined concepts that are its instances at the next lower level).

Information propagates downwards by turning attribute values of the object facet into constant attributes of the class facet (i.e., class attributes whose value is the same for all class instances). For example, Types of vehicle could have a licence type attribute, with value type B for object class B. An attribute with the same name licence type is then a class attribute for class Types of car with a constant value B for all its instances. All subconcepts of class B in the taxonomy similarly have a licence type attribute with value B.

4 MATERIALIZATION

Materialization [PZMY94] is a binary relationship between a class of categories and a class of more concrete objects analyzed in terms of these categories.



Figure 3: An example of materialization.

Figure 3 shows a materialization between classes Types of family car and Family cars. A materialization link is drawn as a line with a “*” on the side of its more concrete class.

Class Types of family car models information that is typically supplied in the catalog of a car dealer, such as model name, sticker price, and available options for the engine size. Class Family cars models information about individual cars, such as manufacture date, serial number, and owner.

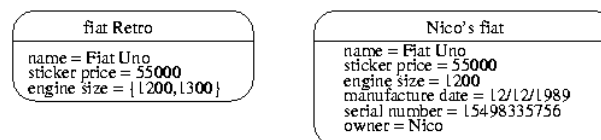


Figure 4: Instances of Types of family car and Family cars of Figure 3.

Figure 4 shows an instance of each class (fiat Retro is an instance of Types of family car and Nico's fiat is an instance of Family cars). The semantics of materialization



expresses that each concrete car (such as Nico's fiat) has exactly one model (fiat Retro), whereas there can be any number of cars of a given model.

The semantics of the abstractness/concreteness relationship also expresses that each car is a concrete realization (or materialization) of a given model, from which it inherits properties in various ways. For example:

- Nico's fiat directly inherits the name and sticker price of its model fiat Retro;
- Nico's fiat has attributes (such as engine size) whose value (1200) is one of the options (1200 or 1300) offered by a multivalued attribute of the same name in object fiat Retro denoting the model of Nico's fiat.

Of course, in addition to the attributes propagated from its model fiat Retro, Nico's fiat has a value for the attributes manufacture date, serial number, and owner of class Family cars.

More detail about the information-propagation mechanisms of materialization can be found in [PZMY94, DPZ02].

5 INTEGRATING CONCEPTS AND POPULATIONS

Figure 5 shows how materialization can realize a correspondence between both views of the taxonomy.

A first type of materialization (like Types of car \rightarrow *Cars labeled (1) in the figure) establishes a systematic bridge between classes of concepts in the intensional view and the corresponding classes of individuals in the extensional view. Other similar materializations include (see Figures 1 and 2): Types of vehicle \rightarrow *Vehicles, Types of truck \rightarrow *Trucks, Bus types \rightarrow *Buses, Types of luxury car \rightarrow *Luxury cars, Types of family car \rightarrow *Family cars, Types of sports car \rightarrow *Sports cars.

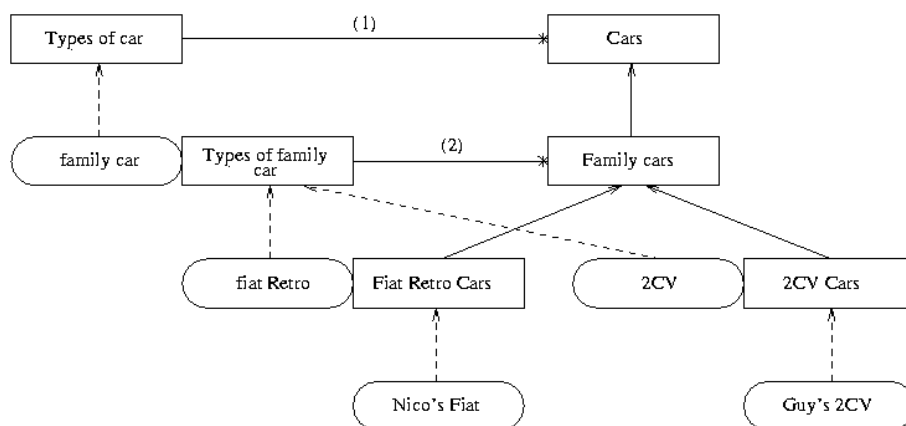


Figure 5: Correspondences via materialization

Figure 5 also shows how both taxonomies merge at their bottom. The materialization **Types of family car** —***Family cars**, labeled (2) in the figure, makes explicit the semantics of the materialization of Figure 3 in terms of two-faceted constructs, and links of classification and generalization.

Object **fiat Retro**, an instance of **Types of family car**, is the object facet of a two-faceted construct whose class facet is class **Fiat Retro cars**, a subclass of **Family cars**, describing all the instances of **Cars** of model **fiat Retro**. **2CV** is another instance of **Types of family car** and **Guy's 2CV** is an instance of its class facet **2CV cars**.

6 SUMMARY

This paper illustrates how the materialization mechanism can establish a systematic correspondence between two taxonomies for the same reality: an “intensional” taxonomy of concepts structured along a class/metaclass dimension and an “extensional” taxonomy of populations structured along a subclass/superclass dimension.

Materialization establishes a bridge at every level of both hierarchies between a class of concepts on the intensional side and a class of individuals of the application domain on the extensional side.

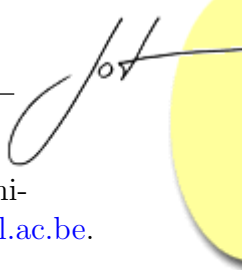
Materialization also establishes a link between both taxonomies at their bottom level.

REFERENCES

- [DPZ02] M. Dahchour, A. Pirotte and E. Zimányi. Materialization and its meta-class implementation. *IEEE Trans. on Knowledge and Data Engineering*, 14(5): 1078-1094, 2002.
- [PM99] A. Pirotte and D. Massart. La matérialisation pour réconcilier deux descriptions des taxinomies. In *Proc. 7è Rencontres de la Société Française de Classification*, Nancy, France, September 1999.
- [PZMY94] A. Pirotte, E. Zimányi, D. Massart, and T. Yakusheva. Materialization: a powerful and ubiquitous abstraction pattern. In J. Bocca, M. Jarke, and C. Zaniolo, editors, *Proc. of the 20th Int. Conf. on Very Large Data Bases, VLDB'94*, pages 630–641, Santiago, Chile, 1994. Morgan Kaufmann.

ABOUT THE AUTHORS

Alain Pirotte is professor in computing science and information at the Université



catholique de Louvain, Louvain-la-Neuve, Belgium, and visiting professor at the Université Libre de Bruxelles, Brussels, Belgium. He can be reached at pirotte@info.ucl.ac.be. See also <http://www.isys.ucl.ac.be/staff/alain/>.

David Massart works as a software engineer at the European Schoolnet Office (<http://www.eun.org/>). He holds a Ph.D. in information science from the Université Libre de Bruxelles, Brussels, Belgium. He also works as an expert for the CEN/ISSS workshop on learning technologies. He can be reached at david.massart@eun.org.