Extended Faceted Ontologies

Yannis Tzitzikas[1,2], Nicolas Spyratos[3], Panos Constantopoulos[1,2], and Anastasia Analyti[2]

[1] Department of Computer Science, University of Crete, Greece
[2] Institute of Computer Science, ICS-FORTH
{tzitzik, panos, analyti}@csi.forth.gr
[3] Laboratoire de Recherche en Informatique, Universite de Paris-Sud, France
spyratos@lri.fr

Abstract. A faceted ontology consists of a set of facets, where each facet consists of a predefined set of terms structured by a subsumption relation. We propose two extensions of faceted ontologies, which allow inferring conjunctions of terms that are valid in the underlying domain. We give a model-theoretic interpretation to these extended faceted ontologies and we provide mechanisms for inferring the valid conjunctions of terms. This inference service can be exploited for preventing errors during the indexing process and for deriving navigation trees that are suitable for browsing. The proposed scheme has several advantages by comparison to the hierarchical classification schemes that are currently used, namely: *conceptual clarity*: it is easier to understand, *compactness*: it takes less space, and *scalability*: the update operations can be formulated easier and be performed more efficiently.

1 Introduction - Motivation

Faceted classification schemes ([1]), hereafter *faceted ontologies*, seem to be superior to hierarchical classification schemes with regard to comprehensibility, storage requirements and scalability ([2]). They are also better suited for indexing collections that are subject to continuous expansion and change.

However, in a faceted ontology *invalid* conjunctions of terms coming from different facets may occur. A conjunction of terms is considered invalid if it cannot be applied to any of the objects of the domain. For example, in a tourist information application, the conjunction `Crete.WinterSports`, where `Crete` belongs to a facet Location and `WinterSports` to a facet Sports (e.g. see Figure 1.(a)), is invalid because there is not enough snow in Crete. In contrast, the conjunction `Crete.SeaSports` is certainly valid. Figure 1.(b) enumerates the set of valid and the set of invalid conjunctions (descriptions) that consist of one term from each facet. The inability to infer the valid conjunctions of terms may give rise to problems in the indexing of the objects (laborious and may erroneous indexing), and in browsing (an invalid conjunction of terms will yield no objects).

Being able to infer the validity of a conjunction in a faceted ontology would be very useful for aiding the indexer and for preventing indexing errors. Such an aid is especially important in cases where the indexing is done by many people. For example, the indexing of Web pages in the Open Directory[1] is done

[1] http://dmoz.org

by more than 20.000 volunteer human editors (indexers). Moreover, if we could infer the valid conjunctions of terms in a faceted ontology then we would be able to generate navigation trees on the fly, which consist of nodes that correspond to valid conjunctions of terms. However, defining manually the set of valid descriptions even for facets of relatively small size, would be a formidable task for the designer.

In this paper we present two extensions of faceted ontologies in the context of which one can infer valid or invalid conjunctions of terms through an inference mechanism. The designer simply declares a small set of valid or invalid conjunctions and other (valid or invalid) conjunctions are then inferred by the proposed mechanism. A full version of this paper can be found in [7].

2 Faceted Ontologies

Roughly, a *faceted ontology* consists of a finite set of facets. Each facet is designed separately, and models a distinct aspect of the domain. For instance, the faceted ontology for the domain of UNIX tools, presented in [2], consists of four facets, namely, "ByAction", "ByObject", "ByDataStructure" and "BySystem". Each *facet* consists of a *terminology*, i.e. a finite set of names or *terms*, structured by a *subsumption* relation. Examples of faceted ontologies can be found in [3], [8], [5], [2]. Below we define precisely what we call ontology, faceted ontology and then we give a semantic interpretation to a faceted ontology.

Def. 1. An *ontology* is a pair $(T, \preceq)$ where T is a *terminology*, i.e. a finite set of names, or *terms*, and $\preceq$ is a *subsumption* relation over T, i.e. a reflexive and transitive relation over T.

Def. 2. A *faceted ontology* is a finite set $\mathcal{F}= \{F_1, ..., F_k\}$ of *ontologies* in which each $F_i = (T_i, \preceq_i)$ is called a *facet*.

Let Obj denote the set of objects of the domain (e.g. the set of all Web pages). Given a terminology T, we call *interpretation* of T over Obj any function $I : T \to 2^{Obj}$.

Def. 3. An interpretation I of T is a *model* of the ontology $(T, \preceq)$ if for all $t, t' \in T$, if $t \preceq t'$ then $I(t) \subseteq I(t')$.

Given a faceted ontology $\mathcal{F}= \{F_1, ..., F_k\}$ let $\mathcal{T}= T_1 \cup ... \cup T_k$ and $\preceq = \preceq_1 \cup ... \cup \preceq_k$. We shall call $(\mathcal{T}, \preceq)$ *the* ontology of $\mathcal{F}$. Now, an interpretation I of $\mathcal{T}$ is a *model* of $\mathcal{F}$, if it is a model of the ontology $(\mathcal{T}, \preceq)$.

Now, a *description* (conjunction) d over $\mathcal{F}$ is either a term $t \in \mathcal{T}$ or a sequence of terms separated by "$\cdot$", i.e. any string derived by the following grammar $d ::= d \cdot t \mid t$. Let D denote the set of all descriptions. An interpretation I of $\mathcal{T}$ can be extended to an interpretation of D as follows: for any description $d = t_1 \cdot t_2 \cdot .. \cdot t_k$ in D we define $I(d) = I(t_1) \cap I(t_2) \cap ... \cap I(t_k)$.

Now, we define semantically what we shall call valid and invalid description. A description d is *valid* in $\mathcal{F}$ if $I(d) \neq \emptyset$ in every model I of $\mathcal{F}$. A description d is *invalid* in $\mathcal{F}$ if $I(d) = \emptyset$ in every model I of $\mathcal{F}$. In the following section we propose two different extensions of an ontology that allows us to infer valid or invalid descriptions from other descriptions that have been declared as valid or invalid by the designer of the faceted ontology.

3 Extended Faceted Ontologies

Def. 4. A *Positive Extended Faceted Ontology*, or *PEFO* for short, is a pair $< \mathcal{F}, P >$ where $\mathcal{F}$ is a faceted ontology and P is a set of descriptions over $\mathcal{T}$. An interpretation I of $\mathcal{T}$ is a *model* of $< \mathcal{F}, P >$ if:
(a) I is a model of $\mathcal{F}$, and
(b) for each $d \in P$, $I(d) \neq \emptyset$.
Now, the set of valid descriptions VD of a *PEFO* $< \mathcal{F}, P >$ is defined as follows: $VD = \{d \in D \mid I(d) \neq \emptyset$ in every model I of $<\mathcal{F}, P>\}$. This means that a description d is valid in $<\mathcal{F},P>$ if there is a description $p \in P$ such that $<\mathcal{F},P> \models p \preceq d$ [2]. If a description is not an element of the set VD, then it is considered invalid (thus we adopt a closed world assumption).

Def. 5. A *Negative Extended Faceted Ontology*, or *NEFO* for short, is a pair $< \mathcal{F}, N >$ where $\mathcal{F}$ is a faceted ontology and N is a set of descriptions over $\mathcal{T}$. An interpretation I of $\mathcal{F}$ is a *model* of $< \mathcal{F}, N >$ if:
(a) I is a model of $\mathcal{F}$, and
(b) for each $d \in N$, $I(d) = \emptyset$.
Now, the set of invalid descriptions ID of a *NEFO* $< \mathcal{F}, N >$ is defined as follows: $ID = \{d \in D \mid I(d) = \emptyset$ in every model I of $<\mathcal{F}, N>\}$. This means that a description d is *invalid* in $<\mathcal{F},N>$ if there is an $n \in N$ such that $<\mathcal{F},N> \models d \preceq n$. If a description is not an element of the set ID, then it is considered valid, i.e. $VD = D \setminus ID$ (thus we adopt a closed world assumption).

Checking description validity requires performing $|P|$ subsumption checks in the case of a *PEFO*, and $|N|$ in the case of a *NEFO*. In both cases, subsumption checking can be performed by the inference mechanism described in [6].

Figures 1.(c) and 1.(d) show how we can specify the valid descriptions of the faceted ontology of Figure 1.(a) (i.e. the sets "Valid Descriptions" and "Invalid Descriptions" as enumerated in Figure 1.(b)) by employing a *PEFO* and a *NEFO* respectively.

The designer can employ a *PEFO* or a *NEFO* depending on the faceted ontology of the application. If the majority of the descriptions are valid then it is better to employ a *NEFO*, so as to specify only the invalid descriptions. Concerning the methodology for defining the set N, it is more efficient for the designer to put in N "short" descriptions that consist of "broad" terms. The reason is that from such descriptions a large number of new invalid descriptions can be inferred. For example in the hypothetical case that we want to specify that all descriptions over the faceted ontology of Figure 1.(a) are invalid, it suffices to put in N one description, i.e. the description `Sports.Location`.

Conversely, if the majority of the descriptions are invalid, then it is better to employ a *PEFO* so as to specify only the valid ones. Concerning the methodology for defining the set P, it is more efficient for the designer to put in P "long" descriptions that consist of "narrow" terms, since from such descriptions a large number of new valid descriptions can be inferred. For example in the hypothetical case that we want to specify that all descriptions over the faceted ontology of Figure 1.(a) are valid, it suffices to put in P just the following description: `SeaSports.WinterSports.Crete.Pilio.Olympus`.

[2] We write $<\mathcal{F},P> \models d \preceq d'$ if $I(d) \subseteq I(d')$ in every model I of $<\mathcal{F},P>$.

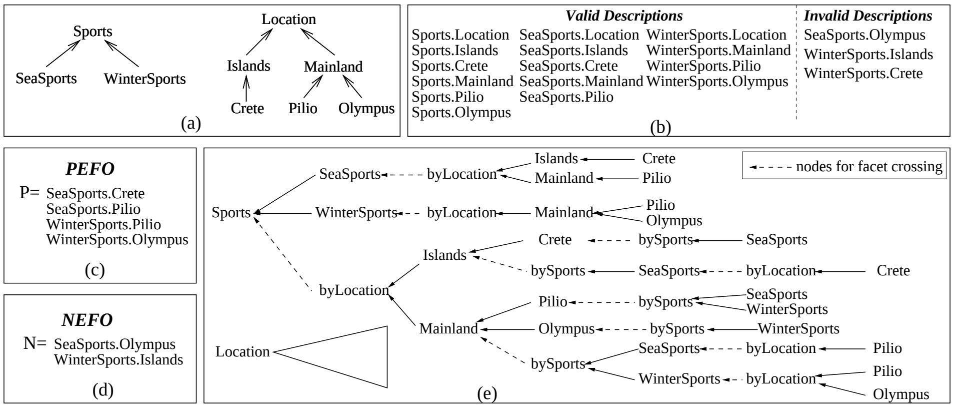

Fig. 1. Our running example

4 Conclusion

We presented a novel approach for indexing and retrieving objects based on multiple aspects or facets. Although even thesauri ([4]) may have facets that group the terms of the thesaurus in classes, our work is original since we described a faceted scheme enriched with a method for specifying the combinations of terms that are valid. Having a *PEFO* or a *NEFO* we can derive *dynamically* a navigation tree such as the one shown in Figure 1.(e), i.e. a tree with nodes that correspond to valid terms only (and nodes for "facet crossing"). Our approach can be used for developing Web catalogs which offer complete navigation trees, require less storage space, and are more comprehensive and scalable.

References

1. Amanda Maple. "Faceted Access: A Review of the Literature", 1995. http://theme.music.indiana.edu/tech_s/mla/facacc.rev.
2. Ruben Prieto-Diaz. "Implementing Faceted Classification for Software Reuse". *Communications of the ACM*, 34(5), 1991.
3. S. R. Ranganathan. "The Colon Classification". In Susan Artandi, editor, *Vol IV of the Rutgers Series on Systems for the Intellectual Organization of Information.* New Brunswick, NJ: Graduate School of Library Science, Rutgers University, 1965.
4. International Organization For Standardization. "Documentation - Guidelines for the establishment and development of monolingual thesauri", 1988. Ref. No ISO 2788-1986.
5. J. H. Sugarman. *"The development of a classification system for information storage and retrieval purposes based upon a model of scientific knowledge generation"*. PhD thesis, School of Education, Boston University, 1981.
6. Yannis Tzitzikas, Nicolas Spyratos, and Panos Constantopoulos. "Deriving Valid Expressions from Ontology Definitions". In *11th European-Japanese Conference on Information Modelling and Knowledge Bases*, Maribor, Slovenia, May 2001.
7. Yannis Tzitzikas, Nicolas Spyratos, and Panos Constantopoulos. "Faceted Ontologies for Web Portals". Technical Report TR-293, Institute of Computer Science-FORTH, October 2001.
8. B. C. Vickery. "Knowledge Representation: A Brief Review". *Journal of Documentation*, 42(3):145–159, 1986.