Faceted Lightweight Ontologies

Fausto Giunchiglia[1], Biswanath Dutta[2], and Vincenzo Maltese[1]

[1] Dipartimento di Ingegneria e Scienza dell'Informazione (DISI)
Università di Trento, Trento, Italy
[2] Documentation Research and Training Centre (DRTC), Indian Statistical Institute (ISI),
8th Mile Mysore Road, Bangalore- 560059, India

Abstract. We concentrate on the use of *ontologies* for the categorization of objects, e.g., photos, books, web pages. *Lightweight ontologies* are ontologies with a tree structure where each node is associated a natural language label. *Faceted lightweight ontologies* are lightweight ontologies where the labels of nodes are organized according to certain predefined patterns which capture different aspects of meaning, i.e., *facets*. We introduce facets based on the Analytico-Synthetic approach, a well established methodology from Library Science which has been successfully used for decades for the classification of books. Faceted lightweight ontologies have a well defined structure and, as such, they are easier to create, to share among users, and they also provide more organized input to semantics based applications, such as semantic search and navigation.

Keywords: Ontologies, Lightweight ontologies, facets, classifications, formal classifications.

1 Introduction

Ontologies are being used in different communities, for different purposes and with different modalities. There are various kinds of ontologies, according to the degree of formality, complexity of the graph structure, and expressivity of the language used to describe them [1]. Ontologies have two main applications. They can be used to *describe* objects or they can be used to *categorize* objects. In this paper, we concentrate on the second use, namely, we are interested in the problem of classifying, e.g., photos, Web pages, books.

Lightweight ontologies are ontologies with a tree structure where each node is associated a natural language label. We sometimes speak of *formal lightweight ontologies* meaning ontologies which can be obtained from lightweight ontologies by translating natural language labels into Description Logics (DL) [12] formulas which capture their meaning ([3] provides an example of how such translation can be done). In formal lightweight ontologies, node formulas stand in the subsumption relation, namely a formula in a node is always more general than the formula in the node below [1, 31]. In the following we talk of lightweight ontologies meaning sometimes formal lightweight ontologies. The context always makes clear what we mean.

A.T. Borgida et al. (Eds.): Mylopoulos Festschrift, LNCS 5600, pp. 36–51, 2009.

Lightweight ontologies allow for automated document classification [1, 16], query answering [1, 21] and also for solving the semantic heterogeneity problem among multiple ontologies [15, 18, 19, 20]. They are definitely a very powerful tool which can be exploited towards the automation of reasoning in data and knowledge management. Still, the adoption of (lightweight) ontologies, so far, has not been as widespread as one would have expected when the work on the Semantic Web started. Among the problems which have been identified are the lack of interest or the difficulties on the user side in building such ontologies [4, 5], but also the fact that ontologies developed for one purpose can hardly being reused for other purposes, or by other users [5].

The goal of this paper is to introduce *faceted lightweight ontologies* as a very promising solution to the problem highlighted above. Faceted lightweight ontologies are defined in terms of *facets*. Recently, facets have been adopted with great success for the design of interfaces to web sites. See, for instance the survey by La Barre [23] and in particular the work done in Flamenco[1] (see for instance [24]), but see also [7,8,9] as an application to knowledge management which is somewhat related, in spirit, to our work. We construct facets following the approach which was first devised by Ranganathan at the beginning of the last century [22] [2] and, in particular, the POPSI Methodology, originally introduced in [26].

Taking the terminology of Library Science, facets are "*aspects of meaning*". They formalize, for any given domain (e.g., medicine, sports, music, science), the main characteristics of that domain and, in particular, the entities or objects which belong to that domain (e.g. in medicine, the body parts), the properties of objects (e.g., in medicine, the various kinds of disease) and the actions which can be taken (e.g., in medicine, surgery or medication). More precisely, a facet is a hierarchy of homogeneous group of terms (nodes), where each term in the hierarchy denotes a primitive atomic concept. Thus we have hierarchies of entities, properties, actions, and so on. We call *background knowledge* [17,14], a *faceted representation scheme,* namely a set of facets that represent the system a-priori knowledge about the domains of interest (see also [13] for an early attempt of defining a faceted representation schema not based on Ranganathan's theory). A faceted representation scheme allows for post-coordination, namely, for constructing complex labels (in Library Science terminology, also called *subjects*) by combining terms from facets at both indexing, classification and searching time. Faceted lightweight ontologies are lightweight ontologies where node labels (formulas) contain only atomic concepts which correspond to primitive concepts taken from the background knowledge.

The rest of the paper is organized as follows. Section 2 introduces and formally defines (classification) lightweight ontologies. Section 3 introduces facets. Section 4 introduces faceted subjects and, then, the notion of faceted lightweight ontology. Finally, Section 5 shows, via an example, how a faceted subject can be constructed according to the POPSI subject indexing system. Section 6 concludes the paper.

---

[1] http://flamenco.berkeley.edu

[2] This theory is widely recognized as a fundamental methodology that guides in the organization of the knowledge in a given domain (see for instance [30]) in terms of basic subjects and relations between them.