A Facet-Based Methodology for Geo-Spatial Modeling

Biswanath Dutta, Fausto Giunchiglia, and Vincenzo Maltese

DISI - Università di Trento, Trento, Italy

Abstract. Space, together with time, is one of the two fundamental dimensions of the universe of knowledge. Geo-spatial ontologies are essential for our shared understanding of the physical universe and to achieve semantic interoperability between people and between software agents. In this paper we propose a methodology and a minimal set of guiding principles, mainly inspired by the faceted approach, to produce high quality ontologies in terms of robustness, extensibility, reusability, compactness and flexibility. We demonstrate - with step by steps examples - that by applying the methodology and those principles we can model the space domain and produce a high quality facet-based large scale geo-spatial ontology comprising entities, entity classes, spatial relations and attributes.

Keywords: space domain, methodology, principle, theory, domain ontology, geo-spatial ontology.

1 Introduction

Space and time are the two fundamental dimensions of the universe of knowledge [12, 3]. The notion of space is essential to understand the physical universe. We consider space as is in accordance with what people commonly understand by this term, which includes the surface of the earth, the space inside it and the space outside it. It comprises the usual geographical concepts, often known as features, like land formations (continents, islands, countries), water formations (oceans, seas, streams) and physiographical concepts (desert, prairie, mountain). It also comprises the areas occupied by a population cluster (city, town, village) and buildings or other man-made structures (school, bank, mine).

Spatial (geo-spatial) and temporal ontologies, because representing a shared understanding of a domain [10], are essential to achieve semantic interoperability between people and between applications. Equally important, the definition of entity types and corresponding properties has become a central issue in data exchange standards where a considerable part of the semantics of data may be carried by the categories that entities are assigned to [20]. As a matter of fact, current standards - for instance the specifications provided for the geographical domain by the Open Geospatial Consortium (OGC)[1] - do not represent an effective solution to the interoperability problem. In fact, they only aim at syntactic agreement [11] by fixing the standard terms and not allowing for variations on the terminology to be used.

[1] `http://www.opengeospatial.org/`

C. Claramunt, S. Levashkin, and M. Bertolotto (Eds.): GeoS 2011, LNCS 6631, pp. 133–150, 2011.
© Springer-Verlag Berlin Heidelberg 2011

Several frameworks have been proposed to build and maintain geo-spatial ontologies [13, 14, 15, 21], and we also recently proposed our multilingual ontology, called GeoWordNet, that overcomes their qualitative and quantitative limitations (as extensively described in [2]). However, to the best of our knowledge no systematic ways, i.e. based on a well founded methodology and guiding principles, for building geo-spatial ontologies have been proposed so far.

Our main contribution is a methodology and a minimal set of guiding principles aimed at modelling the spatial domain and at building the corresponding background knowledge taking into account the classes, the entities, their relations and properties. As explained across this paper, the domain knowledge is organized following the well founded *faceted approach* [3], borrowed from library and information science. Note that the methodology and the guiding principles we propose are not only applicable to the spatial domain, but across domains. In this approach, the analysis of the domain allows the identification of the basic classes of real world objects. They are arranged, per *genus et differentia* (i.e. by looking at their commonalities and their differences), to construct specific ontologies called *facets*, each of them codifying a different aspect of the domain at hand. This allows being much more rigorous in the definition of the domain and its parts, in its maintenance and use [1]. The intended use of this background knowledge is manifold. Identifying the domain specific terminology and corresponding entity names allows using them to annotate, index and search geographical resources as well as for word sense disambiguation.

The rest of the paper is organized as follows. In Section 2 we illustrate our methodology and the guiding principles we propose to model a domain. In Section 3, with some step by step examples, we highlight some of the issues we faced in building the space domain. In Section 4 we describe how we further organize the elements of the domain into three main categories: entity classes, relations and attributes. Section 5 provides some statistics about the space domain, as we modelled it so far. Section 6 concludes the paper and provides our future research directions.

2 The Methodology

Our methodology is mainly inspired by the *faceted approach* proposed by the Indian librarian Ranganathan [3] at the beginning of the last century. In this approach, the domain under examination is decomposed into its basic constituents, each of them denoting a different *aspect of meaning*. Each of these components is called a *facet*. More precisely, a facet is a hierarchy of homogeneous terms, where each term in the hierarchy denotes a primitive atomic concept, i.e. a primitive class of real world objects. In the next two sections we describe the main steps in the creation of the set of facets for a given domain and the guiding principles to be used.

2.1 Steps in the Process

The building process is organized into subsequent phases as follows:

- ***Step 1: Identification of the atomic concepts***. It consists in collecting the terms representing the relevant (according to the purpose) real world entities of the domain at hand. Each term denotes a class of objects. In general, this is mainly