An Ontology Alignment Method for Earth Observations Data Interoperability

Surya S. Durbha, Roger L. King, and Nicolas H. Younan
Center for Advanced Vehicular Systems (CAVS)
Department of Electrical and Computer Engineering (ECE)
Mississippi State University, MS 39762
suryad@cavs.msstate.edu, rking@cavs.mstate.edu, younan@ece.msstate.edu

Abstract-

Earth Observations (EO) are obtained from a multitude of sources and requires tremendous efforts and coordination among different agencies and user groups to come to a shared understanding on a set of concepts involved in a domain. The data and information deluge and the current data archiving and delivery methods need to be transformed into one that allows realization of seamless interoperability. In addition to the several syntactic and structural differences between different data repositories, semantic heterogeneities cause acute data exchange and interoperability problems. The preliminary step to resolve the semantic conflicts is the development of systems that provide formal conceptualization of the local domain entities. This can be achieved through domain specific ontologies which are normally built independent of each other and highly heterogeneous. Currently several organizations are working towards developing domain specific knowledge representations through ontologies. In the oceans domain several programs such as global sea-level observing system, global temperature and salinity profile program, moored and drifting buoys program etc., form part of the Global Ocean Observing System (GOOS) and are responsible to disseminate data on various marine, meteorological and biological parameters.. For instance, the Coastal-Marine Automated Network (C-MAN) [http://www.ndbc.noaa.gov/cman.php] station data typically include barometric pressure, wind direction, speed and gust, and air temperature; however, some C-MAN stations are designed to also measure sea water temperature, water level, waves, relative humidity, precipitation, and visibility [http://www.ndbc.noaa.gov/]. The granularity of parameters served by certain other stations [www.gomoos.org] is much finer (Figure 1(a)).

Similarly in the land cover mapping area the development of thematic land cover data sets is driven by different national or international initiatives. Some of the currently available data products include IGBP DISCOVER, the MODIS land cover product, University of Maryland (UMD) land cover product, GLC 2000, CORINE land cover 1990 and 2000, AFRICOVER.

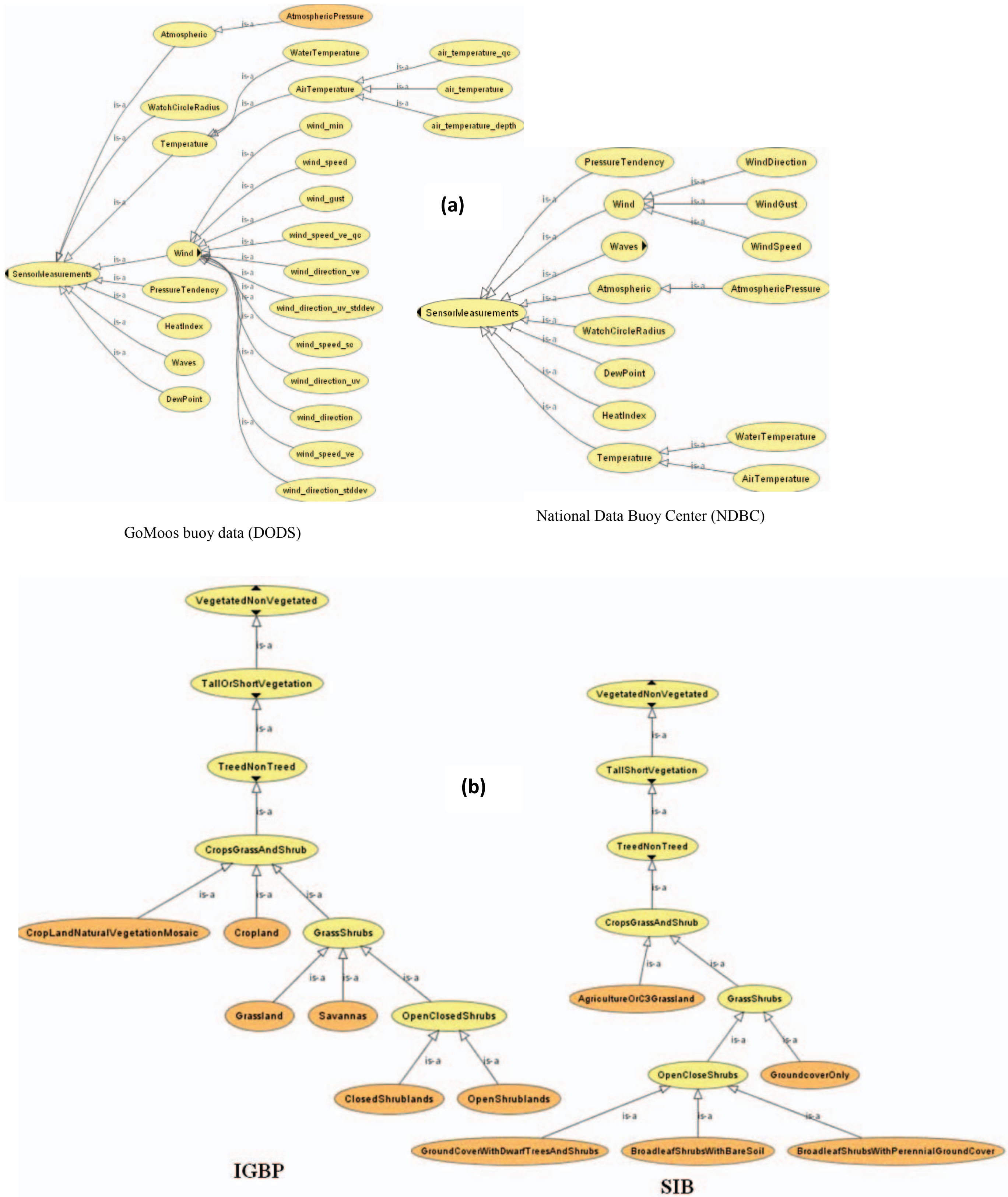

Figure 1. Semantic Conflicts between different information sources (A) Oceans meteorological data served by two different organizations (NDBC and GoMOOS), several conflicts occur that

hinder their interoperability. (B) Conflicts between two land cover classification schemes (IGBP and SiB) in terms of intended meaning of the land cover classes, naming schemes and granularity.

The intent of the different classifications is mainly to reduce the high amounts of information by abstracting from details. However, there is only limited compatibility and comparability between the data sets generated by different organizations. Figure 1(b) shows the two different representations of the land cover in different classification systems. Hence, it is necessary to develop tools that provide alignment between different conceptualizations.

An ontology O is represented as

$$O := (C; HC; RC; HR; I; RI; A) \quad (1)$$

Where C represents Concepts of the schema arranged in a subsumption hierarchy HC, RC represents relations between pairs of concepts arranged in a hierarchy HR, Data is constituted by instances I of specific concepts. These instances are interconnected by relational instances RI [1]. Ontology mapping is the process of finding semantic correspondences between similar elements belonging to different ontologies [2], [3].

The mappings between ontologies provide the means for users to interchange knowledge and thus establish semantic interoperability between different information sources. Ontology mapping enables efficient semantic-based knowledge retrieval from diverse Earth observation data sources.

In our earlier work we proposed a hybrid instance-based algorithm for automated ontology mapping in ocean sensor networks. We employ machine learning techniques (Kernel PCA, support vector machines) and string distance metrics to facilitate the alignment. In this paper, we further refine and generalize the tool to different Earth observations domains to enable the mappings between heterogeneous ontological representations. This work in progress is approaching the problem from a Bayesian perspective and enables kernel-based classification approach by a Gaussian process (GP) model. GP's based classification and regression, has currently gained renewed interest. The earth observations research community has not investigated in depth the applications of GPs as was done for Support Vector Machine (SVM) classifiers. The sparse formulation of the GP models which helps to bring down the complexity of training as well as testing have attracted considerable attention. The GP classifiers are basically Bayesian classifiers derived from Gaussian process priors over functions originally developed for regression [4]. The advantages comes from the explicit probabilistic formulation that yields predictive distributions for test instances and allows standard Bayesian techniques for model selection. GPs allow priors and hyperparameters of the trained models to be easily interpreted and it not only predict the most likely output, but also the probability distribution, which is helpful to combine with other sources of knowledge in the ontologies and metrics such as string distances to make optimal mappings and estimate correspondence similarity measures between concepts in two different ontological representations. We further compare the performance of GP models with the state of the art support vector machines (SVMs).

References:

[1] Patel-Schneider, P.F., Hayes, P., and Horrocks, I. (eds.). OWL Web Ontology Language Semantic and Abstract Syntax. W3C Recommendation 10 February 2004. Latest version available at http://www.w3.org/TR/owl-semantics/for owl.

[2] Doan, A. and Halevy, A. (2005). Semantic integration research in the database community: A brief survey. AI Magazine,

[3] Noy, N. (2004). Semantic integration: A survey of ontology-based approaches. SIGMOD Record.

[4] Rasmussen, C.E. and Williams, C.K.I., Gaussian *Process for Machine Learning*. Cambridge, Massachusetts: The MIT Press, 2006.