An Approach to Facilitate the Integration of Hydrological Data by means of Ontologies and Multilingual Thesauri

Miguel Ángel Latre, Javier Lacasta, Eddy Mojica, Javier Nogueras-Iso, Francisco Javier Zarazaga-Soria

Computer Science and Systems Engineering Department
University of Zaragoza, María de Luna 1, E-50018 Zaragoza, Spain
{latre, jlacasta, eddyma, jnog, javy}@unizar.es
http://iaaa.cps.unizar.es/

Abstract. The general concern about environmental issues has involved the creation of national and international policies that require, at a technical level, the analysis, merging and processing of data obtained from very different sources. This paper proposes an approach for the integration of hydrological data that is based on the use of a multilingual ontology to facilitate the mapping across the local data models in the different sources. The novelty of the proposal is that the multilingual domain ontology is generated automatically by the merging and pruning of existing lexical ontologies. This approach has been tested in the context of the European Water Framework directive for the development of reporting applications in cross-border scenarios. Nevertheless, this approach could be easily extended to other domains.

1 Introduction

The general concern about environmental issues in recent years has involved the creation of national and international policies encouraging the development of information infrastructures to facilitate the cooperative access and exploitation of data coming from different sources from public and private institutions.

An example of this general concern about environmental issues can be found in the European context. The environmental protection is one of the
interests of the European Union and different initiatives and policies in this field are taking place, such as the Water Framework (European Commission, 2000) and INSPIRE (European Commission, 2007) directives.

INSPIRE (INfrastructure for SPatial InfoRmation in Europe) aims at the creation of a European spatial information infrastructure that delivers integrated spatial information services, being environmental information the first application domain tackled by this directive. It is also interesting to take into account that a considerable amount of these environmental initiatives are related to the hydrology domain. The European Water Framework Directive (WFD) is considered to be the most important piece of legislation in this aspect (Usländer, 2005). Its main objective is to achieve an accurate management of all water bodies and reach a “good status” for them by 2015.

This paper proposes an approach for the integration of hydrologic data that aims at discovering implicit relations between hydrologic features that are not usually made explicit in database models. In this particular domain, hydrologists must monitor a great variety of features and phenomena that, although initially disconnected, may affect the status of water bodies. An information retrieval system for this kind of data is presented in this paper.

The approach proposed here is based on the use a multilingual lexical ontology or thesaurus. An ontology is usually defined as “an explicit formal specification of a shared conceptualization” (Gruber, 1993). It is considered as a means for the integration of data because it enables the establishment of a common reference model that facilitates the mapping across the local data models in different sources. Additionally, a domain ontology may help to infer relations that are not usually explicit in the local models and facilitate the combination of different feature types. Although the use of ontologies and thesauri for data integration is not new, the novelty of the proposal is that the multilingual thesaurus focused on the hydrologic domain is automatically generated by the merging and pruning of existing thesauri. The applicability of thesauri for searching and retrieval in digital libraries has promoted the creation and diffusion of well-established thesauri in many different domains. Thus, thesauri can facilitate an important source of information for the development of ontologies focused on specific domains. This automatically generated thesaurus is the main element that allows the information retrieval system presented in this paper to work.

The rest of the paper is organized as follows. Section 2 summarizes the state of the art in ontology based discovery and retrieval. Section 3 describes the information retrieval system this paper is based on, including the methodology for the multilingual thesaurus generation and the results