(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

An approach to facilitate the integration of hydrological data by means of ontologies and multilingual thesauri

M. Á. Latre, J. Lacasta, E. Mojica, J. Nogueras-Iso, and F. J. 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.

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

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

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

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 obtained in the hydrologic domain (3.3). The last section concludes and introduces some ideas on future work.

2 State of the art in ontology based discovery and retrieval

In a Spatial Data Infrastructure context, discovering and accessing suitable geographical information is a crucial task. However, semantic heterogeneity caused by natural language ambiguity (e.g., synonymy, homonymy) makes it difficult to interpret feature property names and user queries.

Some research works have been focused on advancing in the solution for overcoming this heterogeneity. Bernard et al. (2004) describe the architecture of an ontology based discovery and retrieval system of geographical information. In this system, different Web Feature Services are described with metadata which includes a reference to an application ontology that describes the feature types in terms of a shared domain ontology. User queries are processed as follows: users state their queries in terms of the shared domain ontology; then the system expands the user query restrictions with the names of the stored features. Lutz and Klien (2006) work shows the evolution of the previous system. This latter version defines a query language and provides a user interface that helps users to formulate queries.

Other works in this line are the ones proposed by Hübner et al. (2004) and Navarrete (2006). The first one describes an ontology based reasoning system that allows integrating heterogeneous geographical information by resolving structural, syntactic and semantic heterogeneities. The query system supports the specification of queries of the type *concept@location in time*. The user selects a set of registered domain-specific application ontologies (in the thematic, spatial, and temporal domains) based on a

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

common vocabulary and use them to select search terms that are expanded by selecting all equivalences and subconcepts (for the thematic search term), spatially related place names (for the spatial search term), and relevant time periods (for temporal ones). The second one provides a framework to represent semantic relations among the concepts from different datasets of a repository. The system is based on a high level ontology constructed by merging the knowledge provided by the datasets of the repository that describe in a precise and formal way the content of the repository. This ontology is then used to define semantic services or queries that enable agents to find and integrate thematic information. It specifically focuses on finding datasets containing information on a particular theme (including theme subclasses if they are considered of interest); translating the content of a dataset to another compatible vocabulary; and integrating heterogeneous content from different datasets.

Not related to the geo-spatial area but also focused on improving the discovery and retrieval of information using ontologies is the work of Tudhope et al. (2006). It describes a system that uses terminological ontologies (faceted thesauri) to perform query expansion in indexed collections. It describes a semantic closeness algorithm that creates a neighbourhood of semantically related concepts (for retrieval purposes) from a selected one and gives them a weight according to their closeness. Somewhat in the same line is the work of Miles (2006). It analyzes the information retrieval issues caused by the language heterogeneity and proposes a formal theory to describe the ways in which a structured vocabulary may be used to construct and index over a collection of objects. Finally, it compares different expansion techniques in user queries to improve recall and precision.

The information retrieval system presented in this paper goes along the lines of these works, although it presents some differences. We want to combine powerfulness of allowing expert users to make their queries by means of an ontology with the simplicity of offering novice users a search mechanism based on terms from a thesaurus as base for the queries. Additionally, we aim at a system where the data resources are just standard OGC Web Feature Services created and maintained by the appropriate organizations, but where the addition of new information resources to the information retrieval system can be done in a easy and effortlessness way.

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

3 Design and architecture of the system

This section is devoted to present the information retrieval system and the mechanism to generate the multilingual thesaurus. An overview of the system, its functionality, including an example of use, and its architecture are presented below. The generation of the multilingual thesaurus is described next in 3.3, followed by a description of the different components composing the information retrieval system (3.4).

3.1 System overview

The system presented here aims at searching and automatically integrating hydrologic features from different sources on demand, just by providing a hydrology related term or concept. The functionality of the system has been tested in the context of the European Water Framework Directive (WFD) (European Commission, 2000) for the development of reporting applications in cross-border scenarios. This is a similar scenario to the one of the SDIGER (Zarazaga-Soria et al., 2007), an INSPIRE pilot project whose aim was to test the feasibility of developing a cross-border inter-administration SDI to support WFD information access. A web application was developed in order to provide on-line the different map reports about the WFD implementation. In that project, the integration of data coming from French and Spanish data repositories was facilitated by means of ad-hoc software applications applying crosswalks between local data models and common reference models.

The original SDIGER web application aims at using INSPIRE principles for fulfilling the WFD reporting requirements established by the European Commission (European Commission, 2004). The application generates automatically the established reports from data and services belonging to the different WFD Competent Authorities involved in the SDIGER project: The Ebro River Basin Authority *(Confederación Hidrográfica del Ebro,* CHE) and the Adour-Garonne Water Agency *(Agence de l'Eau Adour-Garonne,* AEAG).

This time we want to go a step forward and enable users to search and automatically integrate hydrologic features from different sources on demand. What it is intended is to expand the SDIGER application to be able to work not only with a set of fixed feature types (the SDIGER application worked only with surface water and groundwater bodies) and a fixed set of data sources (the web feature servers of the water agencies involved in the project), but also with other kinds of features, through the

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

use of a hydrology domain ontology and with the help of a multilingual hydrologic thesaurus. The domain ontology we have chosen is built upon the data model proposed by the Common Implementation Strategy of the WFD in the "Guidance document on implementing the GIS elements" (Vogt, 2002). However, an ontology based on the data model proposed by the "Data Specifications" Drafting Team or the Hydrography Thematic Working Group of the INSPIRE directive (INSPIRE Drafting Teams, 2008b, 2008a), once it has been adopted, would be not only equally appropriate in the field of the WFD, but even more generic when applied to the hydrography field.

By making use of the domain ontology, linked with the multilingual thesaurus, the user may request the combination and merging of hydrologic features not necessarily connected in the local data sources due to the use of different modelling approaches, as it is explained in the following sections.

3.2 Architecture

Figure 1 shows the overall architecture of the application. The main component is the *Ontology based IR system*, which takes a term as input and returns a set of features as output. It orchestrates the interactions with the other components of the system: a *Web Ontology Service* (WOS), a *Services Catalog*, a *WFS Query Resolver* and a *WFS Broker*.

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

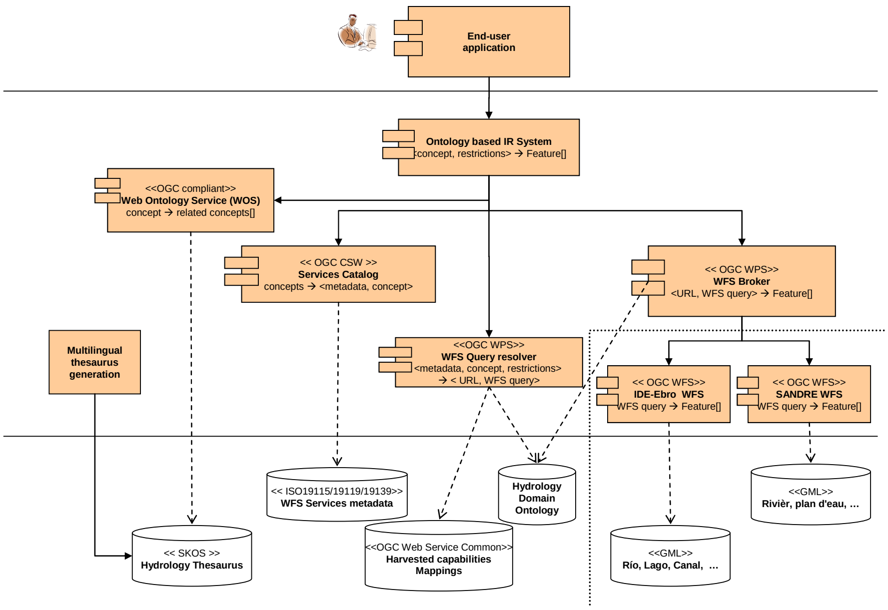

Fig. 1. Architecture of the application for the integration of data

The function of WOS is to expand the search term with related ones, according to the hydrology thesaurus obtained as it is explained in section 3.3. The component is used in order to increase the number of Web Feature Services that are going to be queried for features, and, thus, improve the recall of the information retrieval system. The objective of the query expansion is to solve some of the problems derived in terms of user queries: synonymy, multilinguality and lack of explicit knowledge of the domain (like hierarchical relationships, for instance).

The *Services Catalog* is used to provide WFS instances and feature types that are linked to the term the user is searching for.

The *WFS Query Resolver* purpose is to build the queries that are going to be requested to the services found by the *Services Catalog* according to their local models. It is in charge of selecting the appropriate feature types and translating the restrictions the user may have imposed into a filter encoding query.

Once the appropriate WFS, feature types and, if applicable, the filter encoding query in the local feature models have been obtained, the *WFS Broker* actually performs the queries to the different WFS. It is also in charge of combining the results the system must return.

The relation between the searched concept and the concepts that are related to local repositories is found via the thesaurus.

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

Figure 2 shows a schema that describes how the information retrieval system works. The example shows the particular case of a report about the term “water bodies”, where two different web feature services are part of the system: the WFS of *IDE-Ebro*, the SDI of the Ebro River Basin Authority, and the *SANDRE WFS* of the French Ministry for the Environment. The hydrologic data needed can be accessed through these two servers, located in different data repositories according to the category of the water body (river, lake, transitional or coastal), where different modelling approaches have been used for each server.

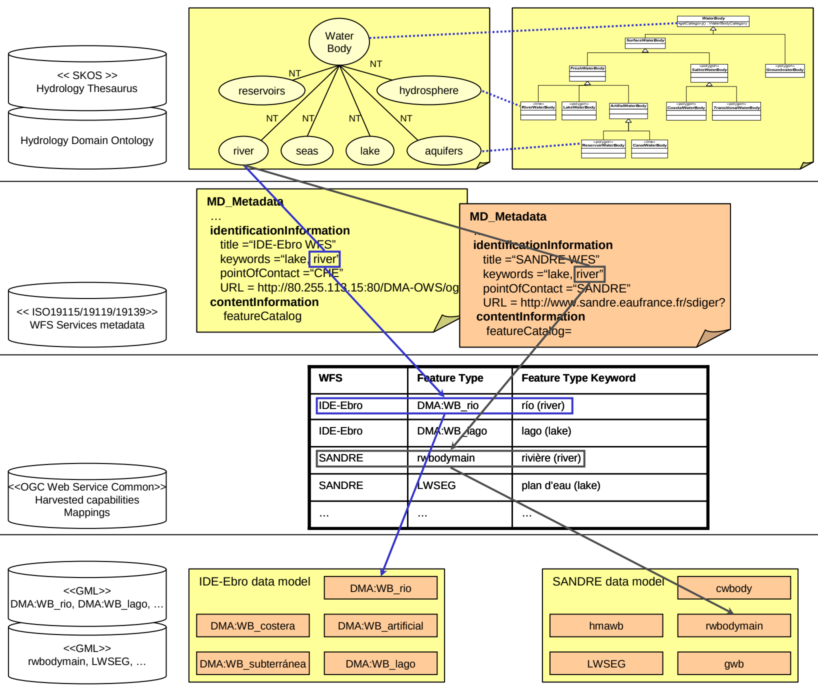

Fig. 2. Example of use

The end-user application makes a request for “water bodies”, which triggers a search of this concept in the hydrologic thesaurus, linked through “narrow-term” relations to the terms “river”, “lake”, “reservoirs”, “seas”, “aquifers” and “hydrosphere” (and their translations to other languages present in the source thesauri). All these concepts but “hydrosphere” appear in the metadata records of the system *Services Catalog*. In the case of “river”, we found that the “DMA:WB_rio” feature type of the *IDE-Ebro WFS* and the “rwbodymain” feature type of the *SANDRE WFS* have a translation of that term as keyword, and the same

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

happens for the rest of the expanded terms ("lake", "reservoirs", "seas" and "aquifers"). Thus, the *WFS Broker* performs ten *getFeature* requests: five to the *IDE-Ebro WFS* to get features of types "DMA:WB_rio", "DMA:WB_lago", "DMA:WB_artificial", "DMA:WB_costera" and "DMA:WB_subterranea", and another five to the *SANDRE WFS* to get features of types "rwbodymain", "LWSEG", "hmawb", "cwbody" and "gwb".

If the results are going to be used for portrayal, the merging component renders them and returns an image to the user application. In the case that a feature collection needs to be returned, the system would propose to the user a mapping between features types and domain ontology concepts. Assuming the domain ontology is the one based on the WFD data model, in the case of the features of type "DMA:WB_rio" returned by the *IDE-Ebro WFS*, the system would propose a map with the concept "riverWaterBody", which the system has deduced because the term "river" of the multilingual thesaurus has been found among the keywords describing the feature type "DMA:WB_rio" in the *IDE-Ebro WFS* capabilities, and since this same term has been previously mapped to the "riverWaterBody" concept of the domain ontology. In addition to this, for each attribute within the features of type "DMA:WB_rio", the system would provide a list of possible matches for mapping with attributes of the concept "riverWaterBody" based on the coincidence of the data type used for its representation, in order to allow the user to map it to the correct one. For instance, the "DMA:WB_rio" attribute "nombre" of type string could be mapped to every string attribute of the concept "riverWaterBody". The user, if no other has done it previously, could select, among all these, the attribute "name" to perform the map. Users could save these mappings in order to be reused by themselves or others, avoiding the need of perform this manual mapping. The user can make queries to the system by either selecting a concept form the multilingual thesaurus, or by choosing a concept of the domain ontology. In the first case, the users are just interested in obtaining features related to a certain term. In the second, they can also provide a set of restrictions (selection of features within a feature type and/or projection of attributes) on the features related to the concept they are interested in, in order to find features satisfying certain conditions.

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

3.3 Generation of a multilingual thesaurus in the hydrology domain

As base for the expansion of queries performed by the system described in this paper, it has been necessary to create a multilingual thesaurus describing the terminology in the hydrology domain. The created terminological model has been based on the list of hydrology related concepts referenced within the European Water Framework Directive. There were two main goals for construction of this thesaurus. On the one hand, we wanted a multilingual resource, while on the other one, we wanted to enrich it with more concepts related to hydrology than the ones described within the Directive text.

The creation process is based on the merging of a set of multilingual terminological ontologies following the thesaurus structure that contain hydrology related concepts with the selected set of hydrologic terms. The output obtained is a multilingual thesaurus specialized in the desired area of knowledge.

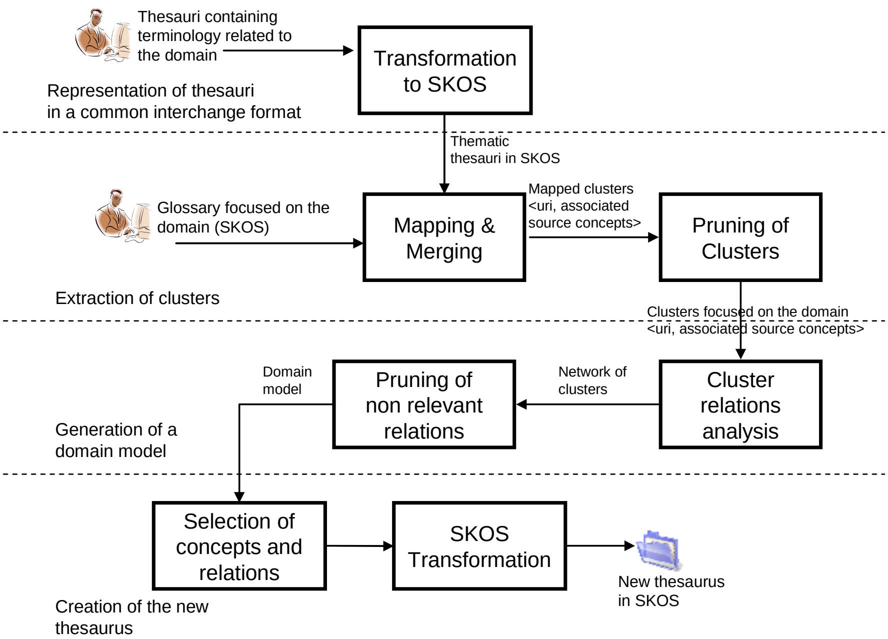

Fig. 3. Work-flow for the generation of a domain specific thesaurus

Figure 3 depicts the different steps of the process, showing the inputs and the produced results. Four different tasks can be highlighted (A

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

detailed description of the three first ones is shown in Lacasta et al. (2007c) work):

- Representation of input thesauri in a common format. This task is devoted to the transformation of the input thesauri into SKOS (Miles et al., 2005) (a W3C initiative for the representation of Knowledge Organization Systems) with the objective of having a homogeneous input to the generation system. Apart from the list of hydrologic terms, the terminological models used as input for the method are the following thesauri: GEMET (the GEneral Multilingual Environmental Thesaurus of the European Environment Agency)[1], AGROVOC (the FAO Agricultural Vocabulary)[2], EUROVOC (the European Vocabulary of the European Communities)[3] and the UNESCO thesaurus[4]. They provide a shared conceptualization in the areas of economics, politics, culture and environment.
- Extraction of clusters. This is the main step and it consists in the detection of intersections between concepts in the different input thesauri, and the set of terms selected in the area of hydrology through the analysis of their lexical similarities. Each set of mapped concepts is grouped into a cluster, which is the name given to a concept in the output thesaurus. A cluster represents a group of equivalent concepts and it is identified with one of the URIs of the original concepts. With the objective of focusing on the hydrology theme, only those clusters that are found to be related to theme are stored. The selection of the clusters is performed with the following criteria:
  - If the cluster contains a concept from the selected European Water Framework Directive list of terms, the cluster is stored.
  - If a concept in the cluster is related to another one in other cluster that fits in the previous case, the cluster is preserved. This is done to add additional terms related to the original concepts of hydrology selected from the Water Framework Directive text.
  - In other case, the cluster is deleted.
- Generation of a domain model. This step consists in connecting the clusters previously extracted. The relations between the concepts assigned to the different clusters are converted into relations between the clusters that contain them. The relations between clusters are labelled with: the types of relations, which are derived from the original

[1] http://www.eionet.europa.eu/gemet
[2] http://www.fao.org/aims/ag intro.htm
[3] http://europa.eu/eurovoc/
[4] http://www.ulcc.ac.uk/unesco/

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

types of relations between concepts; and a weight that represents the number of occurrences for each original relation type between the concepts of the inter-related clusters. Additionally, it is possible to reduce the size of the resulting network by selecting only those concepts related between them by relations of at least a minimum weight. For example, it is possible to obtain a reduced network with only those clusters that have relations of at least weight 3 (that is, they have been found in three of the original thesauri).

- Generation of a new thematic thesaurus. The last step of the defined process is to transform the network of clusters into a thesaurus. The generation of the thesaurus consists in taking the clusters of the network and organizing them into a hierarchical model. The clusters are transformed into concepts of the new thesaurus; one of the labels of the original concepts within the cluster is selected as preferred label. With respect to the thesaurus structure, each relation is marked with the type that has more occurrences. Additionally, those concepts that do not have broader relationship are marked as top terms. Finally, the generated structure is reviewed to verify that the BT/NT relationships structure does not contains cycles. If a cycle is found, it is removed by replacing the NT/BT relationship that generates the cycle by a *related* relationship.

The result thesaurus has been generated using the complete network of concepts as base (all the generated concepts and relations). It contains 322 concepts with 966 preferred labels and 2424 alternative labels. With respect to the relations, the number of *broader*/*narrower* pairs is 239 and the number of *related* relationships is 203. Figure 4 shows a subset of the generated thesaurus using the ThManager tool[5] (Lacasta et al., 2007b). The figure shows a branch of the thesaurus starting from the "land cover" concept. It shows the generated hierarchy containing the different types of water bodies.

[5] ThManager is an OpenSource tool for the creation and visualization of terminological ontologies stored in SKOS format (see http://thmanager.sourceforge.net/).

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

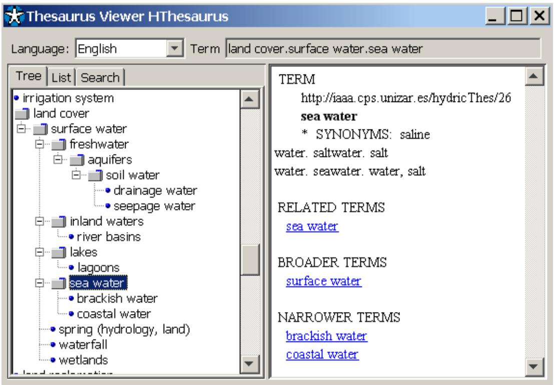

Fig. 4. Visualization of a part of the generated hydrologic thesaurus

3.4 Components of the system

3.4.1 Web Ontology Service

The *Web Ontology Service* is an OGC Web Services Architecture specification compliant component whose purpose is to facilitate the management and use ontologies and thesauri to other web components requiring them. Designed as a centralized service, the architecture of this service aims at reducing the cost of creation of a new ontology or thesaurus, improving reusability and avoiding duplicities and inconsistencies. The architecture and a detailed revision of the functionality of these services are described in Lacasta et al. (2007a). In the context of this paper, this service is used to solve some of the problems derived from the ambiguity of user queries (e.g., synonymy, multilinguality, or lack of explicit knowledge of the domain). This is done by expanding the user search terms with other related ones, according to the multilingual thesaurus obtained in the previous section. This reformulation of the queries has as final objective to increase the number of WFS to be queried for features, and, thus, improve the overall recall of the information retrieval system.

3.4.2 Services Catalog

The function of the *Services Catalog* is to provide to the system with WFS that are linked to the term the user is searching for. It is a standard OGC

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

Services Catalog that access ISO19115/19119/19139 metadata (Nogueras-Iso et al., 2009), which have been created automatically through a crosswalk from the different capabilities XML files the Web Feature Services included in the system return to the *getCapabilities* request. This approach has two main advantages. On the one hand, in order to include a new WFS into the system, just its URL and its capabilities files are needed. On the other hand, the capabilities XML files include a list of the feature types, together with a list of keywords for each feature type served by a particular WFS. If the capabilities file is thorough enough, the transformed metadata record can be used perfectly to find appropriate WFS related to a hydrologic concept, avoiding the need of manually editing and completing them. Since we are working with an thesaurus built from a set of multilingual thesauri, the language in which the keywords are written in the capabilities is irrelevant (provided that the language is supported by the original thesauri).

3.4.3 WFS Query Resolver

The *WFS Query Resolver* purpose is to make up the actual queries that are going to be requested to the Feature Services found by the *Services Catalog*. It is in charge of selecting the appropriate feature type to query the WFS. In order to do so, it requests to the WFS its capabilities and finds out which feature type is linked with the keyword or concept the user is interested in searching.

Additionally, it is also in charge of translating the restrictions users may have imposed. This set of restrictions on the features they are interested in must have been established by using the domain ontology, and the *WFS Query Resolver* translates them into the appropriate filter encoding queries. In order to do this, the user restrictions terms (attributes and, possibly, values) are translated into the appropriate ones for a particular WFS. In this case, the data model of any WFS that can be accessed by our information retrieval system must be carefully analyzed, through the metadata provided by *getCapabilities* and *describeFeatureType* requests (and it is even probable that an additional detail of information about the data model would be needed). Then, a manual mapping between elements of the WFS feature model and the domain ontology must be done. Not only feature types must be mapped to the domain ontology concepts, but also their attributes and their domain values must be mapped (Fallahi et al., 2008). In the case of quantitative attributes, it could be necessary to identify measurement units, since conversion of units could be necessary. And, in the case of qualitative attributes, a mapping between the possible values the attributes can have in the feature model and in the domain

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

ontology must be also performed if these values are a controlled list or a thesaurus. An OGC Feature Catalog would have been used if it has been possible, but there is no possibility of storing keywords of each feature type and attribute, and it does not allow establishing mappings between feature types of different collections.

The approach where users provide just a hydrologic term to perform the search is less powerful, but it is definitely much more simple from the use and maintenance points of view: users do not have to make a query using a domain ontology (they can freely use any term to perform the search), and the mapping between the user request and the WFS feature models can be done automatically and on-the-fly (just WFS capabilities information is needed to perform the mapping). Thus, the scalability of this approach is enormous. In the second approach, where users must express their searches in terms of a domain ontology, a previous work of mapping feature types, attributes and domain values must have been done. Anyway, this work of maintenance in order to add new WFS can be alleviated by getting the users to do it, as it is explained in the next section.

3.4.4 WFS Broker

Once the appropriate Web Feature Services, feature types and, if applicable, the filter encoding query of the local feature models have been obtained, the *WFS Broker* actually performs the actual queries to the different services. The *WFS Broker* is also responsible of combining the results the system must return:

- The simplest way of combining the results consists in generating a map with the spatial data of the returned features and, then, returning an image instead of a feature collection. Obviously, since the returned features are provided in the form of a map, the only use of this returning mechanism is the portrayal of them. In this case, it can be considered that the *WFS Broker* is acting like a Web Map Service with Style Layer Descriptor capabilities.
- Combining the results into a single GML file and into a unique feature type. In order to do that, the system would provide the user with a mapping to the domain ontology. Since there is a map between the term in the multilingual thesaurus and one of the keywords that appear describing the feature type in each of the capabilities of the WFS that have been queried, and that the terms of the thesaurus and the domain ontology concepts have already mapped, the system can automatically propose mappings between the different feature types and the domain ontology. Furthermore, the system can also propose syntactically correct

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

mappings between the different attributes of the feature type and the ones of the domain ontology concept, based on equivalence of data types of the attributes. Users could select the correct mappings between attributes, and, this way, make mappings between the local models of the WFS and the domain ontology. These mappings can be reused, by the same user in posterior queries, and by the system itself, to be used by the *WFS Query Resolver* when transforming complex user queries into the appropriate filter encoding queries, as it has been mentioned previously. In this last case, the mappings should be validated by the system administrator, since the may be used by queries of other users of the system.

At this moment, we have implemented the first approach, and we are working on the second one.

4 Conclusions

This paper has presented an information retrieval system that facilitates the integration of hydrologic data and the discovery of implicit relations between features, not usually found directly in local data repositories. The relation between the searched concept and the concepts that are related to local repositories is found via a multilingual thesaurus generated starting from a set of thesauri from different knowledge areas and a selected list of terms focused on the domain. The system takes as input a search term or query, uses the multilingual thesaurus to expand it, locates and queries appropriate Web Feature Services, and returns the results as a map or as a feature collection.

Further work will be devoted to the study of some inferred relations and how they can contribute to the improvement of hydrologic models, to the returning mechanism of data as a feature collection and how to orchestrate the process as a service chain using formal languages for service composition such as BPEL (Business Process Execution Language).

Acknowledgments

This work has been supported by the research project TIN2006-13301, granted by the Spanish Ministry of Education and Science, by the Spanish Ministry of Science and Innovation through the projects TIN2007-65341 and PET2006_0760 from the National Plan for Scientific Research, Development and Technology Innovation and Zeta Amaltea, S.L. The

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

work of Eddy Mojica has been partially supported by the University of Zaragoza, Banco Santander and Funcación Carolina through the program “Ayudas para realizar estudios de doctorado dirigidas a estudiantes latinoamericanos. Curso 2008–2009.”

References

Bernard, L., Einspanier, U., Haubrock, S., Hübner, S., Klien, E., Kuhn, W., Lessing, R., Lutz, M., and Visser, U. (2004). Ontology-based discovery and retrieval of geographical information in spatial data. Geotechnology Science Report 4, Institute of Geoinformatics, Münster.

European Commission (2000). Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for Community action in the field of water policy. Official Journal of the European Union.

European Commission (2004). Common Implementation Strategy for the WFD. Reporting sheets for 2005 reporting. Version 5.0. Technical report, European Commission - DG Environment D.2.

European Commission (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Official Journal of the European Union.

Fallahi, G. R., Frank, A. U., Mesgari, M. S., and Rajabifard, A. (2008). An ontological structure for semantic interoperability of GIS and environmental modeling. *International Journal of Applied Earth Observation and Geoinformation*, 10(3):342–357.

Gruber, T. (1993). A translation approach to portable ontology specifications. *ACM Knowledge Acquisition, Special issue: Current issues in knowledge modeling*, 5, Issue 2(KSL 92-71):199–220.

Hübner, S., Spittel, R., Visser, U., and Vogele, T. (2004). Ontology-based search for interactive digital maps. *IEEE Intelligent Systems*, 19(3):80 – 86.

INSPIRE Drafting Teams (2008a). D2.8.I.8 INSPIRE Data Specification on Hydrography – Draft Guidelines. Technical report, European Commission.

INSPIRE Drafting Teams (2008b). Drafting Team “Data Specifications” — deliverable D2.3: Definition of Annex Themes and Scope. Technical report, European Commission.

Lacasta, J., Nogueras-Iso, J., Béjar, R., Muro-Medrano, P., and Zarazaga-Soria, F. (2007a). A web ontology service to facilitate interoperability within a spatial data infrastructure: applicability to discovery. *Data & Knowledge Engineering*, 63(3):947–971.

Lacasta, J., Nogueras-Iso, J., López-Pellicer, F. J., Medrano, P. M., and Zarazaga-Soria, F. J. (2007b). ThManager: An open source tool for creating and visualizing SKOS. *Information Technology and Libraries (ITAL)*, 26(3):39–51.

(Draft) Lecture Notes in Geoinformation and Cartography (LNG&C). Advances in GIScience. 2009, p. 155-171. ISSN 1863-2246.

Lacasta, J., Nogueras-Iso, J., Zarazaga-Soria, F., and Muro-Medrano, P. (2007c). Generating an urban domain ontology through the merging of a cross-domain lexical ontologies. In *Proceedings of 2nd Workshop of COST Action C21 - Towntology: Ontologies for urban development: conceptual models for practitioners*, pages 60–74.

Lutz, M. and Klien, E. (2006). Ontology-Based Retrieval of Geographic Information. *Journal of Geographical Information Science*, 20(3):233–260.

Miles, A. (2006). Retrieval and the semantic web. A theory of retrieval using structured vocabularies. Master's thesis, Oxford Brookes University.

Miles, A., Matthews, B., and Wilson, M. (2005). SKOS Core: Simple Knowledge organization for the WEB. In *Proc of Int Conf on Dublin Core and Metadata Applications*, pages 5–13, Madrid, Spain.

Navarrete, A. (2006). Semantic integration of thematic geographic information in a multimedia context. PhD thesis, Pompeu Fabra University.

Nogueras-Iso, J., Barrera, J., Rodríguez, A. F., Recio, R., Laborda, C., and Zarazaga-Soria, F. (2009). *Spatial Data Infrastructure Convergence: Research, Emerging Trends, and Critical Assessment*, chapter Development and deployment of a services catalog in compliance with the INSPIRE metadata implementing rules. The Netherlands Geodetic Commission (NGC), The Netherlands. accepted for publication.

Tudhope, D., Binding, C., Blocks, D., and Cunliffe, D. (2006). Query expansion via conceptual distance in thesaurus indexed collections. *Journal of Documentation*, 62(4):509–533.

Usländer, T. (2005). Trends of environmental information systems in the context of the European Water Framework directive. *Environmental Modelling & Software*, 20(12):1532–1542.

Vogt, J. (2002). Guidance document on implementing the GIS elements of the Water Framework Directive. Technical report, Commission of the European Communities.

Zarazaga-Soria, F. J., Nogueras-Iso, J., Latre, M. A., Rodríguez, A., López, E., Vivas, P., and Muro-Medrano, P. (2007). *Research and Theory in Advancing Spatial Data Infrastructure Concepts*, chapter Providing SDI Services in a Cross-Border Scenario: the SDIGER Project Use Case, pages 107–119. ESRI Press.