A Practical Example of Semantic Interoperability of Large-Scale Topographic Databases Using Semantic Web Technologies

Koen Aerts[1,2], Karel Maesen[1] and Anton Van Rompaey[2]

[1] K.U.Leuven, R&D Divisie SADL (Spatial Applications Division Leuven)
Celestijnenlaan 200E, B-3000 Leuven, Belgium, Tel: +32 16 32 97 32, Fax: +32 16 32 97 24
[2] Physical and Regional Geography Research Group – Katholieke Universiteit Leuven
Celestijnenlaan 200E, B-3000 Leuven, Belgium, Tel: +32 16 32 64 33, Fax: +32 16 32 29 80
[Koen.Aerts, Karel.Maesen]@sadl.kuleuven.be, Anton.VanRompaey@geo.kuleuven.be

SUMMARY

This paper describes the development of an Ontology-Driven Geographic Information System for the integration of vector-based, large-scale topographic data, within the scope of the WalkOnWeb project. This research focuses on solving semantic heterogeneity of topographic feature classes. The use of new semantic web technologies, such as the Web Ontology Language, is investigated and demonstrated in a practical example of road networks in topographic databases. The proposed methodology is based on a Formal Concept Analysis to extract classifying properties from topographic feature classes. In the next step ontologies are developed describing the topographic feature classes and their classifying properties in terms of an upper ontology. The ontologies enable automatic reasoning about topographic feature classes based on their semantic meaning. The ability to derive super/sub class relationships and equivalences between topographic feature classes is an important development in order to achieve semantic interoperability.

KEYWORDS: *Semantic Web, Semantic Interoperability, Ontologies, Topographic Databases, Description Logic*

INTRODUCTION

Topographic maps supply a general image of the Earth's surface and the geographic features on it, such as roads, rivers, houses, vegetation, relief, etc. Topographic maps have a wide variety of applications, ranging from guiding a hiker through the landscape, to strategic decision support for policy makers, or even supporting emergency services in case of a crisis. Since the digital revolution, however, a whole new world of applications has become possible. Topographic data is no longer only available in paper format but also in vector-based topographic databases, allowing for more complicated applications.

An important condition for such complex applications is the ability to share and integrate information from different sources. Most of the European countries already have a vector-based, large-scale (1:10.000) topographic database for their territory. Although these databases contain similar data, merging data from different databases is not a straightforward task. Several kinds of heterogeneity need to be solved. Some of them already have well known solutions, such as transformations between different coordinate and projection systems, others, such as semantic heterogeneity, still require further investigation.

This paper focuses on the specific problem of the semantic heterogeneity in topographic databases. The first section describes how heterogeneity arises in topographic databases through the process of conceptualisation. The different kinds of heterogeneity that can be distinguished in topographic databases are listed in the second section. In the third section a new framework to resolve semantic heterogeneity using semantic web technologies such as the Web Ontology Language (OWL) is proposed. The semantic interoperability problems of transnational hiking path cartography are taken as an example application. The last section gives a preview of ongoing research.

This research is done within the scope of the WalkOnWeb project[1], which aims to develop a new electronic publishing model for hiking information across Europe. A topographic map is an indispensable aid for the hiker on tour. Therefore, WalkOnWeb develops an interoperable geographic information system (GIS) for large-scale topographic databases. Both an online internet version to plan the walk and a mobile personal digital assistant (PDA) version for use during the walk are being developed.

AN ABSTRACTION PARADIGM FOR THE GEOGRAPHIC WORLD

It is common practice in cartography to use an object-oriented approach as the basic conceptualisation of space (Fonseca, 2001), especially in topographic maps. The object model represents the world as a surface occupied by discrete, identifiable features with a geometrical representation (point, line or polygon) and descriptive properties.

Fonseca (2001) and Fonseca et al (2002) describe a five-universe paradigm for the abstraction of geographic features from the geographical world. The five universes are the physical universe, the cognitive universe, the logical universe, the representation universe, and the implementation universe. The physical universe is the real world with everything that people are capable of perceiving. A geographic phenomenon in the real world is captured by the cognitive system of a person and is classified and stored in the human mind. The representation of the real world object in the human cognitive system is done within the cognitive universe. The assignment of concepts to objects of the physical world takes place by agreement of a community that shares common perceptions, for example a national mapping agency (NMA).

Topographic concepts can be defined as abstract specifications of real world objects as they exist in a specific community or national mapping agency. Concepts can have properties, which are the attributes or characteristics that can be assigned to a specific concept. Properties distinguish the concepts they characterise. Topographic concepts are organised in logical frameworks. The formalisation of the conceptualisations of the world gives us explicit formal structures: the ontologies that are part of the logical universe. These formal descriptions of topographic concepts will serve as a basis for interoperability between topographic databases from different national mapping agencies that all have a different conceptualisation of the real world.

HETEROGENEITY IN TOPOGRAPHIC DATABASES

There exists a great heterogeneity between the objects that are distinguished in topographic databases in Europe. National topographic databases are usually maintained by the National Mapping Agency (NMA) of the respective country. Due to historical, organisational and technical constraints, NMA’s have a different conceptualisation of the world, resulting in heterogeneity of topographic features. As a consequence, topographic data sets are not interoperable because of different heterogeneities. Bishr (1998) distinguishes three types of heterogeneity: 1) semantic, 2) schematic and 3) syntactic. The method proposed in this paper mainly focuses on semantic heterogeneity. In the following sections the three kinds of heterogeneity are briefly explained.

Semantic heterogeneity

Semantic heterogeneity means that a topographic feature can have more than one description or interpretation due to different conceptualizations of the respective NMA’s (transition between the physical universe and the cognitive universe). Bishr distinguish two types of semantic heterogeneity: cognitive and naming heterogeneity.

[1] http://www.walkonweb.org

Cognitive heterogeneity occurs when two communities have a different view about the same real word object. For instance: the road transportation community sees a road as a feature that has a function in transportation and has topological connectivity properties. Hydrologists, on the other hand, see roads as surfaces with particular water permeability and run-off properties.

Naming heterogeneity means that concepts which are semantically equivalent and refer to the same real world phenomena, might be named differently. For instance, “watercourse” and “river” might be two names for the same concept.

Schematic heterogeneity

Schematic heterogeneity occurs because the same logical or conceptual schema can be implemented in different ways in the physical schema. In a relational database, for example, a table can be created for each concept separately; alternatively a table may correspond to several concepts – the concepts being distinguished be means of a type column. Other typical sources of schematic heterogeneity are the choice of data type for attributes.

Syntactic heterogeneity

A final type of heterogeneity is due to the differences in the data format. The same data (physical model) can be stored as a set of shape files or as a GML document.

USING ONTOLOGIES FOR SEMANTIC INTEROPERABILTY

Semantic interoperability can be defined in different ways. In this research the following definition is used: “the problem of semantic interoperability is the problem of ensuring that exchanged information is processed consistent with its intended meaning”. The intended meaning is most often not explicit stated in the database structure, but is captured in manuals or specifications of the database. Or it may even just be available in the working knowledge of the community.

It is the goal of this research to make the knowledge about the intended meaning available for the system in such a way that the system can reason about and work with the meaning of the concepts. For instance the system should be able to derive equivalences between topographic concepts from different topographic databases or find super/sub class relations between them.

The strategy presented in this paper makes use of semantic web technologies such as ontologies and the Web Ontology Language (OWL). The following sections explain the concept of an ontology and how it is used to overcome semantic interoperability. The last section clarifies the approach using the practical example of road networks.

Ontologies

A commonly used definition of ontology in the computer science community was established by Thomas Gruber (1993): “An ontology is an explicit specification of a conceptualisation”. An ontology is a formally described, machine-readable collection of terms and their relationships expressed in an ontology language. The Web Ontology Language (OWL) (W3C, 2004; Lacy, 2005) is an ontology language especially developed for the semantic web and influenced by the theory of Description Logics.

In order to overcome the semantic heterogeneity in different topographic databases, we develop ontologies for these databases. The ontologies contain all the topographic feature classes that are known in the database in a hierarchical structure. Each of the feature classes is defined in terms of an upper ontology. An upper ontology is a very general ontology that attempts to describe very general concepts such space, time, matter, object, event, action, etc., which are independent of a particular problem or domain (Guarino, 1998). Several upper ontologies already exist, such as the Suggested

Upper Merged Ontology[2] (SUMO) or OpenCYC[3]. These upper ontologies are very extensive and currently not available in the correct format (OWL). For the time being we have opted to construct our own, provisional upper ontology, containing only a limited vocabulary to express the meaning of topographic feature classes.

Semantic definitions of topographic feature classes

A semantic definition of a topographic feature class consists of a set of necessary and sufficient conditions that a topographic feature must fulfil in order to be an instance of that class. This means that every individual topographic feature that belongs to a specific feature class has these properties, and conversely that a feature satisfying all conditions is an instance of the class.

The necessary and sufficient conditions are deduced from the definitions of the topographic features of the NMA's by a Formal Concept Analysis (Wille, 1992). Kokla and Kavouras (2001, 2002) proposed this method for the formalisation and integration of geographic categorisations. In our approach, the Formal Concept Analysis is used as a heuristic to extract necessary and sufficient properties from topographic concepts and possibly to classify the topographic concepts into a hierarchical structure. The Formal Concept Analysis will only be used for property extraction. The classification will be done based on the ontology.

When applying the Formal Concept Analysis, we distinguish between "classifying properties" and "non-classifying" properties. Classifying properties are those properties that distinguish one topographic feature from another. For example: it is a classifying property of a motorway that it has separated carriageways. In contrast, the fact that a motorway has a width, does not distinguish it between other kinds of roads, so it is not a classifying property. The distinction between classifying and non-classifying properties is not always clear. For example: for roads, the property width is non-classifying because it does not differentiate between roads, but in contrast, watercourses can be divided into categories based on their width, making width a classifying property.

The implementation of these concepts into a workable GIS requires translation of the object properties into a computer language. After assigning classifying properties to feature classes in the Formal Concept Analysis, these properties are translated into necessary and sufficient conditions in OWL, using the Protégé tool[4] with the OWL Plugin (Knublauch et al, 2003). These conditions are expressed in terms of our own upper ontology. The great advantage of defining topographic feature classes in OWL is the ability to reason about the classes. A reasoner, such as Pellet[5], can be used to infer additional facts about classes that were not explicitly stated in the ontology. Super and sub class relations or equivalences between topographic feature classes of different topographic databases can be inferred using their semantic description. This implementation will be explained using the road network data from a French topographic database and the Digital Geographic Information Exchange Standard (DIGEST, 2000).

An example of the road networks

Road networks are a good example of similar topographic features that have different categorisations according to different NMA's. The most general feature in the road network is a road, defined as "An open way maintained for vehicular use" (DIGEST, 2000). Most topographic databases know this general feature class "Road", but at lower level, sub classes or categorisations of roads, many differences occur. By defining the semantic meaning of different kind of roads in an ontology, a reasoner will be able to find relationships and equivalences between different categorisations of roads.

---

[2] http://www.ontologyportal.org/
[3] http://www.opencyc.org/
[4] http://protege.stanford.edu/index.html
[5] http://www.mindswap.org/2003/pellet/index.shtml

The starting point is a list of natural language descriptions of the topographic feature classes under investigation. For example: a common definition of a “Motorway” on topographic maps is: “a road with separated carriageways, grade separated crossings and only accessible for high-speed motor traffic”. In French topographic maps an “Autoroutier” is defined as “a road with separated carriageways and grade separated crossings for at least 5 km” (IGN, 2002).

The first step now is the Formal Concept Analysis to extract the classifying properties for both concepts. They are both kinds of roads, so they inherit all the classifying properties of “Road”. Furthermore the fact that they have separated carriageways and grade separated crossing, distinguishes them from other kinds of roads. For motorways also the restriction to high-speed motor traffic is a classifying property. For “autoroutier” it is not explicitly stated to only have high-speed motor traffic, but from experience or feedback from the database specialist, it can be derived that this is also valid for “autoroutier”.

The next step is to translate the classifying properties into necessary and sufficient conditions in our ontology. The two concepts are defined in their own respective database ontology, identified by a unique namespace. The general top-level concepts and properties are defined in the upper ontology, with namespace UO. Figure 1 shows the conditions we assert to a motorway. Under the “Necessary and sufficient” header the translated classifying properties are listed. Under the “Inherited” header, properties are listed that the feature inherits from its super classes, in this case from “Road”. Non-classifying properties can be inserted, as an option, under the “Necessary” header

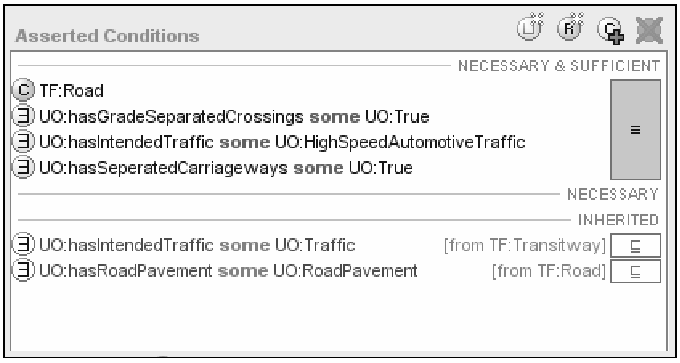

Figure 1: Protégé screenshot of the asserted conditions of a motorway

After specifying the necessary and sufficient conditions for the topographic feature classes, a reasoner can be used to derive super/sub class relationships or equivalences between the road categorisations. Concepts are stated to be equivalent if they have exactly the same necessary and sufficient conditions. A concept is a subclass of another as it has at least the same necessary and sufficient conditions as its super class and possibly some additional necessary and sufficient conditions.

Figure 2 shows the asserted ontology structure of the road network of the French NMA and of a general road network. The asserted ontology structure is explicitly stated by the ontology designer. Figure 3 shows the integrated ontology structure as classified by a reasoner. Equivalences such as the one between “Motorway” and “Autoroutier” are connected by a double arrow (they are both super and sub classes of each other). Further categorisations are shown with a single arrow. For example a “Route Principale” from the French topographic database can be further divided into “Primary and secondary roads” from the generic categorisations.

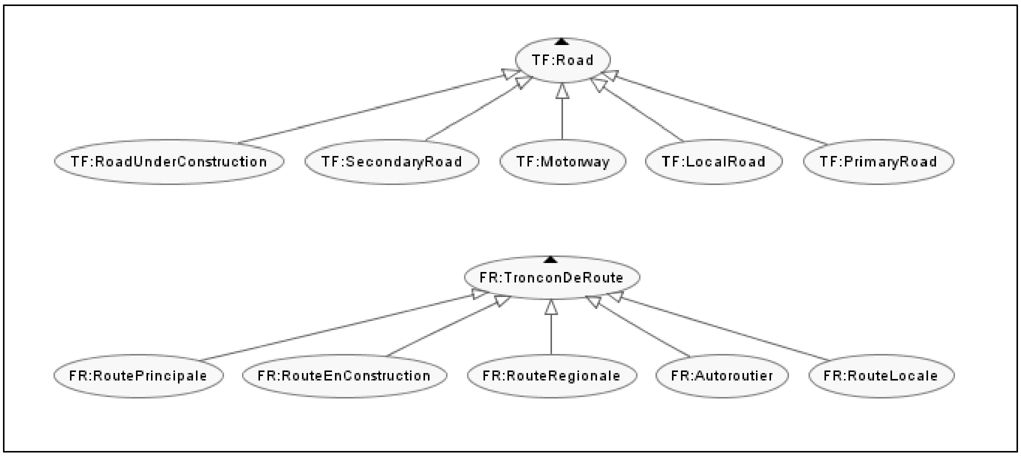

Figure 2: Asserted (before classifying) ontologies of road networks, TF = generic topographic feature, FR = French topographic feature

In order to derive as most relations as possible between topographic concepts, the upper ontology should have a restricted vocabulary. Often the same properties for a topographic concept can be expressed in different ways. For example: “has separated carriageways” is equivalent to “has a central reservation”. When not explicitly stated that these properties are the same, a reasoner can not deduce that roads X with separated carriageways and road Y with a central reservation are semantically identical. By keeping the upper ontology as limited as possible and obliging the user to only extend it if no other property is present to express the meaning of the concept, the chance of missing relationships by using different expressions is limited.

ONGOING RESEARCH

This paper shows a working example of semantic interoperability at a conceptual level. The binding with the database structure is made through annotation properties. Ongoing research is focussing on the database level. How can semantic interoperability be achieved for database attributes? For example: in two different topographic databases we have an attribute ‘road width’. At first sight equivalence can be assumed. After a closer look however, the first road width is only the measurement of the passable part of the lane, while the second also takes into account the width of the sidewalks. How to deal with differences in measuring units?

An important limitation of using OWL for semantic interoperability is the expressiveness of the ontology language concerning data type constructs. Statements such as “a watercourse is called a river when its width is at least 25 meter” are not possible in OWL at the moment. These kinds of constructs are however frequently used in the geographic community. The W3C Semantic Web Best Practices and Development Working Group has set up a task force to address this issue. Pan and Horrocks (2006) propose a decidable extension to OWL-DL, named OWL-Eu, which supports customised datatypes.

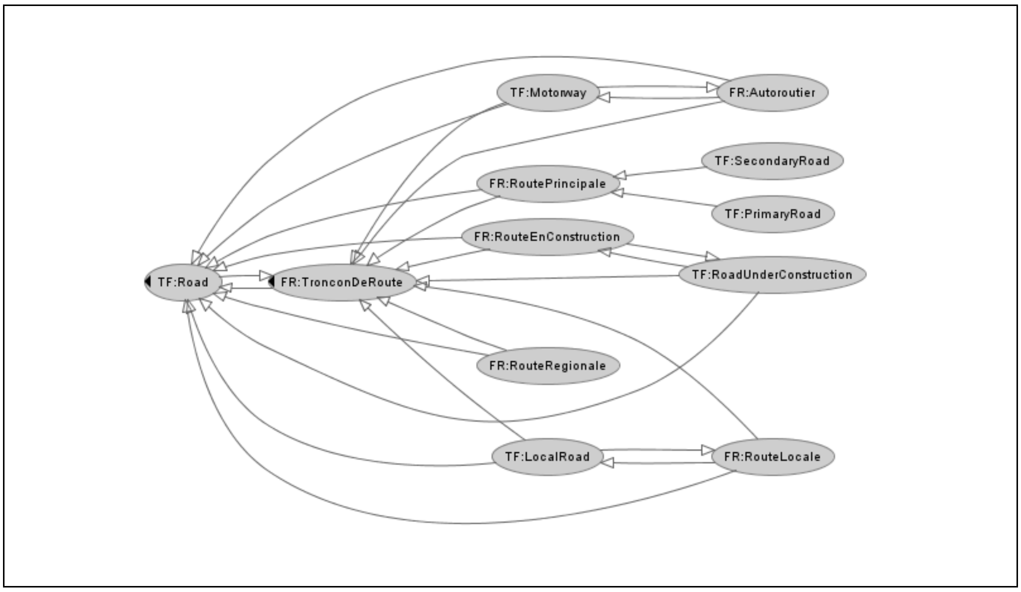

Figure 3: Automatic classification of road networks, TF = general topographic feature, FR = French topographic feature

CONCLUSIONS

In this research the methodological foundations were developed for the semantic integration of vector-based, large-scale topographic databases. It is shown that the use of new semantic web technologies such as ontologies and OWL enables reasoning about the meaning of topographic feature classes. This meaning can be explored by Formal Concept Analysis and translated into an ontology. Given explicit definitions of topographic feature classes in terms of an upper ontology, it is possible to derive super/sub class relationships and equivalences between topographic feature classes. This is an important development in order to achieve semantic interoperability. At this moment semantic interoperability is achieved on a conceptual level. Ongoing research is presently focussing on the database level.

The next step is to implement this technology into a Geographic Information System. In the scope of the WalkOnWeb project an ontology-driven GIS will be developed to demonstrate the usability of the technology. This system will provide interoperability between topographic databases for the field of cross-border hiking. This should enable a hiker to choose different rendering styles of the topographic data according to his wishes. Another aim is the language independent description of a hiking trail in terms of topographic features. This will allow the hiker to always use his familiar terminology for topographic features, even if this terminology is not explicitly stored in the topographic database.

ACKNOWLEDGEMENTS

The authors would like to thank the IST programme, Framework Programme 6 from the European Commission for funding the WalkOnWeb project.

BIBLIOGRAPHY

Bishr, Y., 1998. Overcoming the Semantic and Other Barriers to GIS Interoperability. International Journal of Geographical Information Science, 12(4):299-314.

DIGEST, 2000. The Digital Geographic Information Exchange Standard (DIGEST), Edition 2.1, Part 4, Annex A: Feature Codes. Digital Geographic Information Working Group (DGIWG), 48 p.

Fonseca, F.T., 2001. Ontology-Driven Geographic Information Systems. PhD-thesis, University of Maine, USA, 118 p.

Fonseca, F.T., Egenhofer, M., Davis, C., Câmara, G., 2002. Semantic Granularity in Ontology-Driven Geographic Information Systems. Annals of Mathematics and Artificial Intelligence, 36(1 –2): 121-151.

Gruber, T., 1993. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2): 199-220.

Guarino, N., 1998. Formal Ontology and Information Systems. In Guarino, N. (ed.): Formal Ontology in Information Systems. Proceedings of FOIS'98, IOS Press, Amsterdam, pp. 3-15.

IGN, 2002. BD Topo Pays, Version 1.2, Descriptif de Contenu, Version 1.0. Institut Géographique National, Paris, France, 118 p.

Knublauch, H., Fergerson, R.W., Noy, N.F. and Musen, M.A., 2003. The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. In 3rd International Semantic Web Conference (ISWC 2004), Hiroshima, Japan, 2004.

Kokla, M. and Kavouras, M., 2001. Fusion of Top-level and Geographical Domain Ontologies based on Context Formation and Complementarity. International Journal of Geographical Information Science, 15(7): 679-687.

Kokla, M. and Kavouras, M., 2002. A Method for the Formalization and Integration of Geographical Categorizations. International Journal of Geographical Information Science, 16(5): 439-453.

Lacy, L.W., 2005. OWL: Representing Information Using the Web Ontology Language. Trafford Publishing, Victoria, BC, Canada, 282 p.

Pan, J.Z. and Horrocks, I., 2006. OWL-Eu: Adding Customised Datatypes into OWL. Journal of Web Semantics, 4(1): 29-39.

Wille, R., 1992. Concept Lattices and Conceptual Knowledge Systems. Computers and Mathematics with Applications, 23(6-9): 493-515.

W3C, 2004. OWL Web Ontology Language, Overview. W3C Recommendation. http://www.w3.org/TR/owl-features (accessed 12/12/2005)