Geographic Ontologies: Survey and Challenges

Robert Laurini (Lyon, France) & Okba Kazar (Biskra, Algeria)
< robert.laurini@insa-lyon.fr > | < kazarokba@yahoo.fr >

The goal of this paper is to present some concepts concerning geographic ontologies and especially the use of spatial relations. Ontologies can be defined as a kind of semantic networks for the description of the real world; they are essentially graphs between concepts linked by relations such as is_a, has_a, part_whole. But the scope of geographic ontologies is to describe not only the geographic features, but also their spatial relationships. Usually, only topological relations are defined, but other spatial and geographic relations must be considered as well. After a short presentation of ontologies in general, the characteristics of geographic ontologies will be detailed. Finally, some future challenges shall be outlined.

Keywords: ontologies, geographic ontologies, geographic knowledge, geographic reasoning, spatial relationship

1. Introduction

In this paper, the general concept of ontology is presented, followed by a more detailed discussion of geographic ontologies. Those ontologies not only model conventional geographic features with their semantic relations, but also regarding spatial relations between those features. This paper aims to identify these spatial relations, and to show how to use them for modeling and manipulating geographic ontologies.

Etymologically, "geography" means the description of the Earth, while "ontology" refers on the discourse about existing things. Hence, "geographic ontology" means the description of things existing on the Earth, i.e. of geographic features.

For decades, ontologies have been used in information technologies to describe knowledge in a domain as a kind of semantic networks, especially for the interoperability of databases and for knowledge description in artificial intelligence.

But the management of space requires additional concepts; this is the reason why geographic ontologies show several peculiarities.

This paper will be organized as follows: After this short introduction, the characteristics of geographic ontologies will be discussed. Finally, some future challenges shall be explained.

2. Generalities about ontologies

Ontologies provide a solution for information search on the web, which has two problems, namely silence and noise. Silence means that existing information can not be accessed, while noise referes to the reception of undesired or unwanted information. The main problem of this approach comes from syntax. On the other hand, for excellent interoperability of various databases, a semantic approach has proven to be the ideal solution, i.e. based on the use of the ontology.

2.1. Role and definition

Ontologies were developed within artificial intelligence (AI) research to facilitate sharing and reuse of knowledge. Since the early 1990s, ontologies have become a popular research topic investigated by several AI research communities, including knowledge engineering, natural-language processing and knowledge representation. More recently, the topic of ontologies also receives increasing attention in fields such as intelligent information integration, cooperative information systems, information retrieval, electronic commerce, and knowledge management. A main reason, ontologies are becoming so popular is that they provide a shared and common understanding of some domains, which can be communicated between people and application systems. Ontologies are developed to provide machine-processable semantics of information sources that can be communicated between different agents (software and humans).

Various definitions of ontologies have been proposed during the last decades. A description that, in our opinion, offers a very appropriate explanation of what is the essence of an ontology can be found in Gruber (1993): "An ontology is a formal, explicit specification of a shared conceptualization". 'Conceptualization' refers to an abstract model of some phenomenon in the world, which identifies the relevant concepts of that phenomenon. 'Explicit' means that the type of concepts used and the constraints on use are explicitly defined. 'Formal' refers to the fact that the ontology should be machine-readable, and 'shared' means that several actors must archive a consensus.

2.2. Ontology types

Depending on the level of generality, different ontology types may be identified that fulfill different roles in the process of building a knowledge base system (KBS), (Guarino 1998, van Heijst et al. 1997). Among others, we can distinguish the following ontology types:

- *Domain ontologies* capture the knowledge valid for a particular type of domain (e.g. electronic, medical, mechanic, digital domain).
- *Metadata ontologies* like *Dublin Core* (Weibel et al., 1995) provide a vocabulary to describe the content of on-line information sources.
- *Generic* or *common sense ontologies* aim at capturing general knowledge about the world, providing basic notions and concepts for e.g. time, space, state, event etc. (Fridman-Noy & Hafner, 1997). As a consequence, they are valid across several domains. For example, an ontology about mereology (part-of relations) is applicable in many technical domains (Borst & Akkermans, 1997).
- *Other types* of ontologies are so-called *method* and *task ontologies* (Fensel & Groenboom, 1997; Studer et al., 1996). Task ontologies provide specific terms for particular tasks (e.g. 'hypothesis' belongs to the diagnosis task ontology), and method ontologies provide terms specific to particular propose-and-revise method ontologies (PSM), (e.g. 'correct state' belongs to PSM). Task and method ontologies provide a reasoning point of view on domain knowledge.

2.3. Ontology examples

Prominent ontology examples are WordNet[1], CYC[2], TOVE3[3], and W3C[4]. WordNet (cf. Fellbaum, 1999) is an online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives and

[1] https://wordnet.princeton.edu/
[2] http://psych.utoronto.ca/users/reingold/courses/ai/cyc.html
[3] http://www.eil.utoronto.ca/theory/enterprise-modelling/tove/
[4] https://www.w3.org/TR/owl-features/

adverbs are organized in synonym sets, each representing one underlying lexical concept. Different relations link these synonym sets. Developed by the Cognitive Science Laboratory at Princeton University, WordNet contains around 100,000 words organized in a taxonomy.

WordNet groups words into five categories: noun, verb, adjective, adverb, and function word. Within each category, words are organized by concepts (i.e.: word meanings) and via semantic relationship between words.

Examples of these relationships are:

- *Synonymy*: Similarity in meanings of words, used to build concepts represented by a set of words.
- *Antonymy*: Dichotomy in meaning of words - mainly used for organizing adjectives and adverbs.
- *Hyponymy*: Is-a relationships between concepts. This is-a hierarchy ensures the inheritance of properties from superconcepts to subconcepts.
- *Meronymy*: Part-of relationships between concepts.
- *Morphological relations* which are used to reduce word forms.

2.4. Ontology Components

An ontology consists of several components. The names of these components depend on the expressivity of the ontology (or, in general, of the knowledge representation language) used. Despite this, core components are (in large parts) shared between different ontologies. The main components of ontologies are: concepts, instances, and relations (Lord, 2010), see Figure 1.

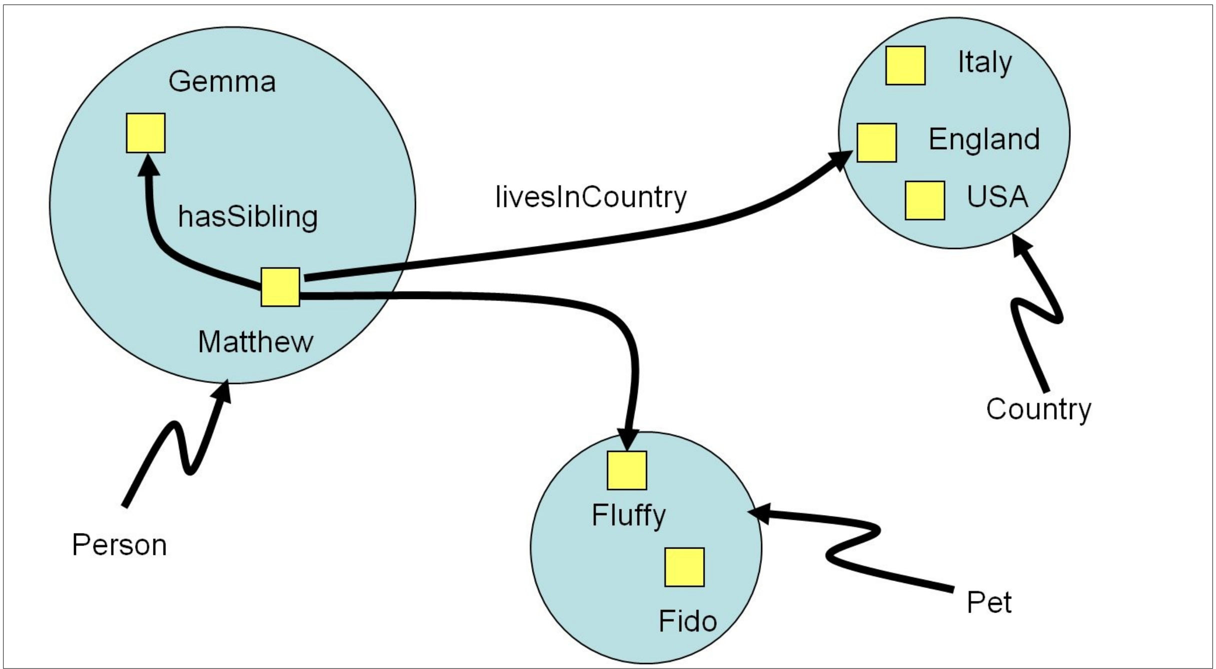


Figure 1: Example of an ontology

2.4.1. Concepts

Concepts, also called classes or types, are a core component of most ontologies. A concept represents a group of different individuals that share common characteristics, which may be more or less specific. For instance, *person* is a concept that represents a set of individuals (persons). One concept may be a sub-concept (also known as subclass, or kind) of another concept; this means that, given concept C' is a sub-concept of C, then any individual of C' will also be an individual of C. Concepts may also share relationships with each other; these relationships describe the way individuals of one concept relate to the individuals of another.

2.4.2. Instances

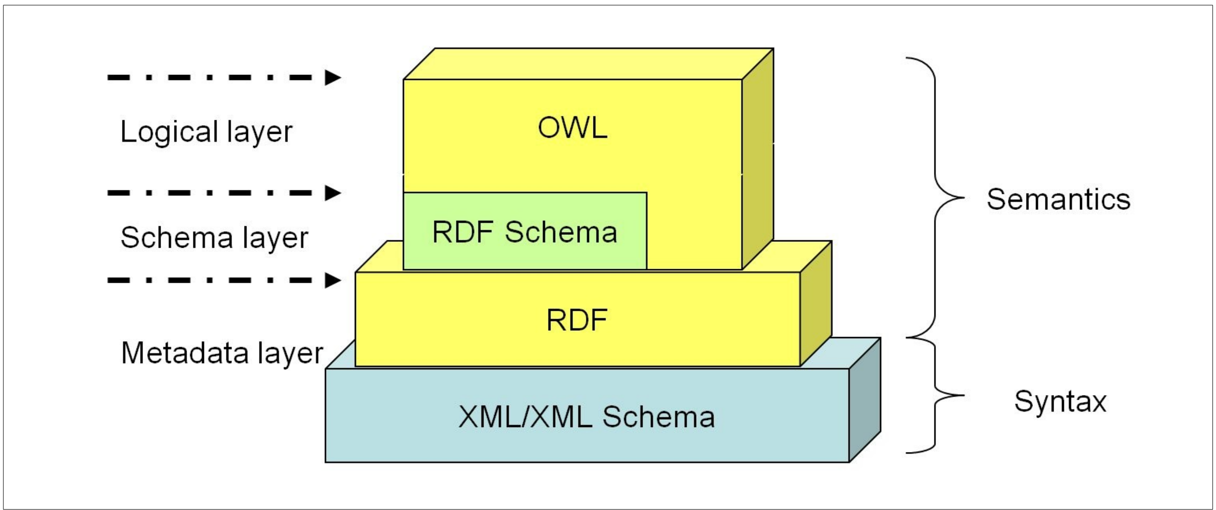
Individuals, also known as instances or particulars, are the base unit of an ontology. They are the things that the ontology describes or potentially could describe. Individuals may model concrete objects such as people or machines; they may also model more abstract objects such as countries, a person's job or a function.

2.4.3. Relations

Relations within an ontology describe the ways individuals relate to each other. Relations normally can be expressed directly between individuals, e.g. the relation hasSibling might link the individuals Matthew and Gemma; or between concepts, e.g. the relation livesInCountry might link the concept Person with the concept Country. In the latter case, a relationship between all individuals of the concepts is being described.

2.5. Ontology Languages

Literature offers a variety of description languages to express ontologies, based on different representations (Figure 2). According to Gómez-Pérez & Corcho (2002), some of them are based on XML syntax, such as Ontology Exchange Language (XOL[5]), Resource Description Framework (RDF) and RDF Schema, and OWL[6] (Web Ontology Language). All of those languages are created by the World Wide Web Consortium (W3C) working groups (Figure 2).



Figures 2: Ontology languages (Tao, 2013).

OWL (Smith, Welty, & McGuinness, 2004) is a key to the semantic web that was proposed by the Web Ontology Working Group of W3C. It is a language extension of the RDF Schema, representing a general-purpose ontology language that contains all necessary constructors to formally describe most of the information management definitions: classes and properties, with hierarchies, and also range and domain restrictions.

5 https://www.sri.com/work/publications/xol-xml-based-ontology-exchange-language
6 http://www.w3.org/2001/sw/wiki/OWL

Basic OWL has the power to express richer properties :

- Symmetric properties (If A connects B then B also connects A)
- Transitive properties (If A is contained in B and B is contained in C then A is contained in C)
- Functional properties (A property that has at most one value for each object)
- Inverse properties (If A is related to B in a way of relation X, then B is related to A in a way of relation Y. X and Y relations are inversed.

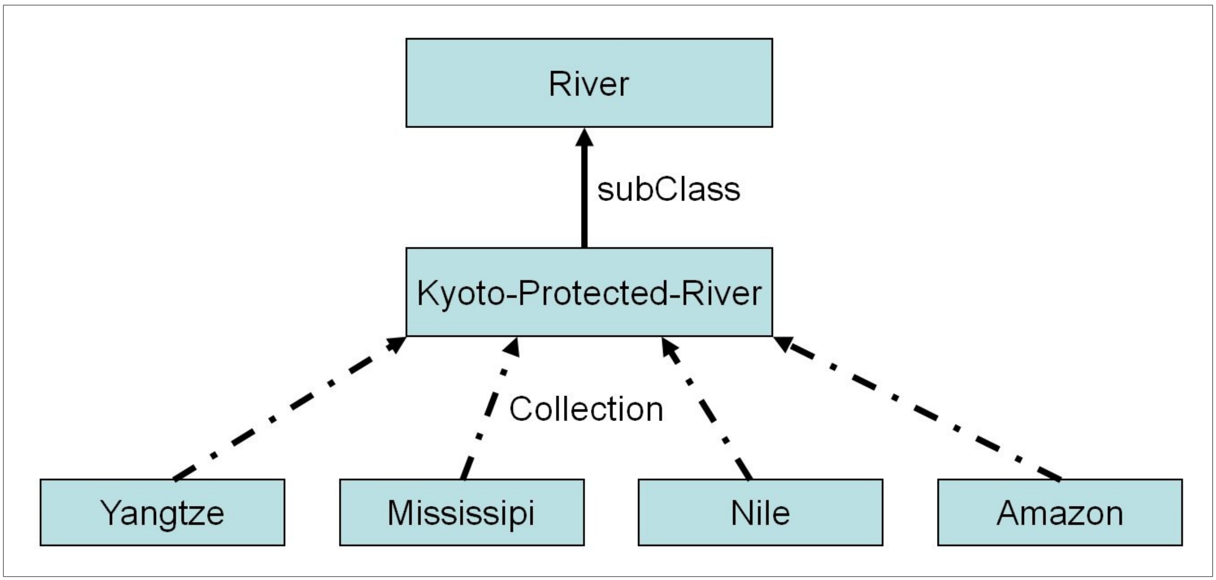


Figure 3. Example of an ontology described in the text with OWL

Example of an OWL class definition 5 (Figure 3):

```
<owl:Class rdf:ID="Kyoto-Protected-River">
   <rdfs:subClassOf rdf:resource="#River"/>
         <owl:oneOf rdf:parseType="Collection">
         <geo:River rdf:about="http://www.china.org/rivers#Yangtze"/>
         <geo:River rdf:about="http://www.us.org/rivers#Mississippi"/>
         <geo:River rdf:about="http://www.africa.org/rivers#Nile"/>
         <geo:River rdf:about="http://www.s-america.org/rivers#Amazon"/>
   </owl:oneOf>
</owl:Class>
```

Restrictions in property definitions:

- `onProperty`: Specifies on which property the restriction will be applied.
- `allValuesFrom`: Specifies which values are accepted by the property.
- `hasValue`: Specifies which value the property has to be exactly.
- `someValue`: Specifies that the property must have at least a value.
- `cardinality`: Specifies the occurrence of the property.
- `minCardinality`: Specifies the minimum occurrence of the property.
- `maxCardinality`: Specifies the maximum occurrence of the property

To illustrate some issues of classifying geographic objects, Table 1 (issued from Kavouras, 2005) depicts, how some existing systems describe water bodies.

Table 1: Different classifications of water bodies according to Kavouas (200()

Ontology	Category_type
CORINE Land Cover	Peat bog Water course Water body
MEGRIN	Bog Canal Lake / pond Salt marsh Salt pan Watercourse
WordNet	Body of water Bog Canal Lake Pond Salt pan Watercourse

5. Characteristics of Geographic Ontologies

In the past, geographic ontologies organized geographic objects with conventional relations. In Figure 4, the beginning of such an ontology is shown. However, it can be seen immediately that such vision is insufficient to describe space. From the different issues relevant to geographic ontologies, just a few shall be mentioned subsequently, namely, the status of space, the spatial relations and linguistic problems. An additional example can be taken from Sowa (2009), where a prototypic geographic ontology is described via geometric types of its features. But now, geographic ontologies integrate better representations of space and spatial relationships.

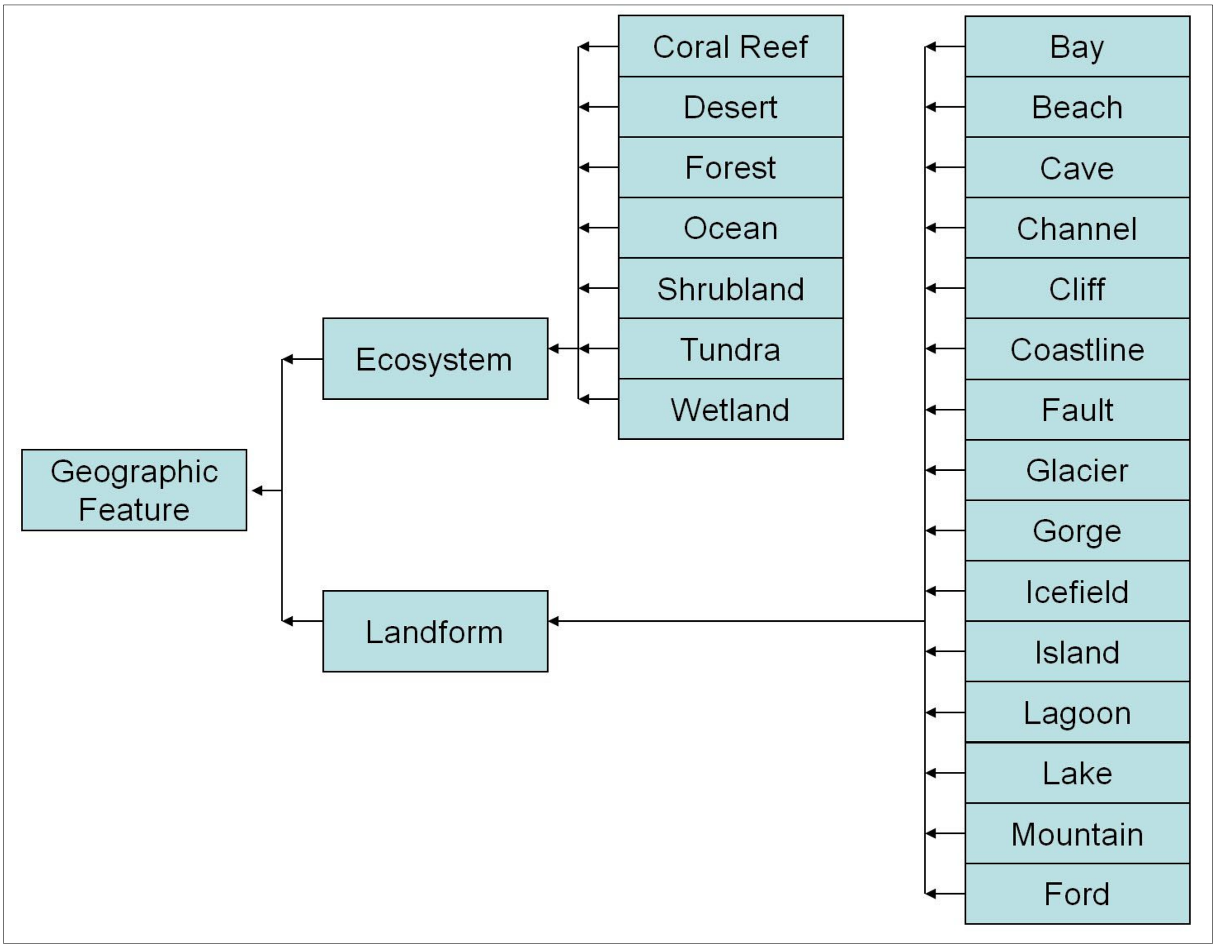


Figure 4. Example of a geographic ontology only using is_a relations.

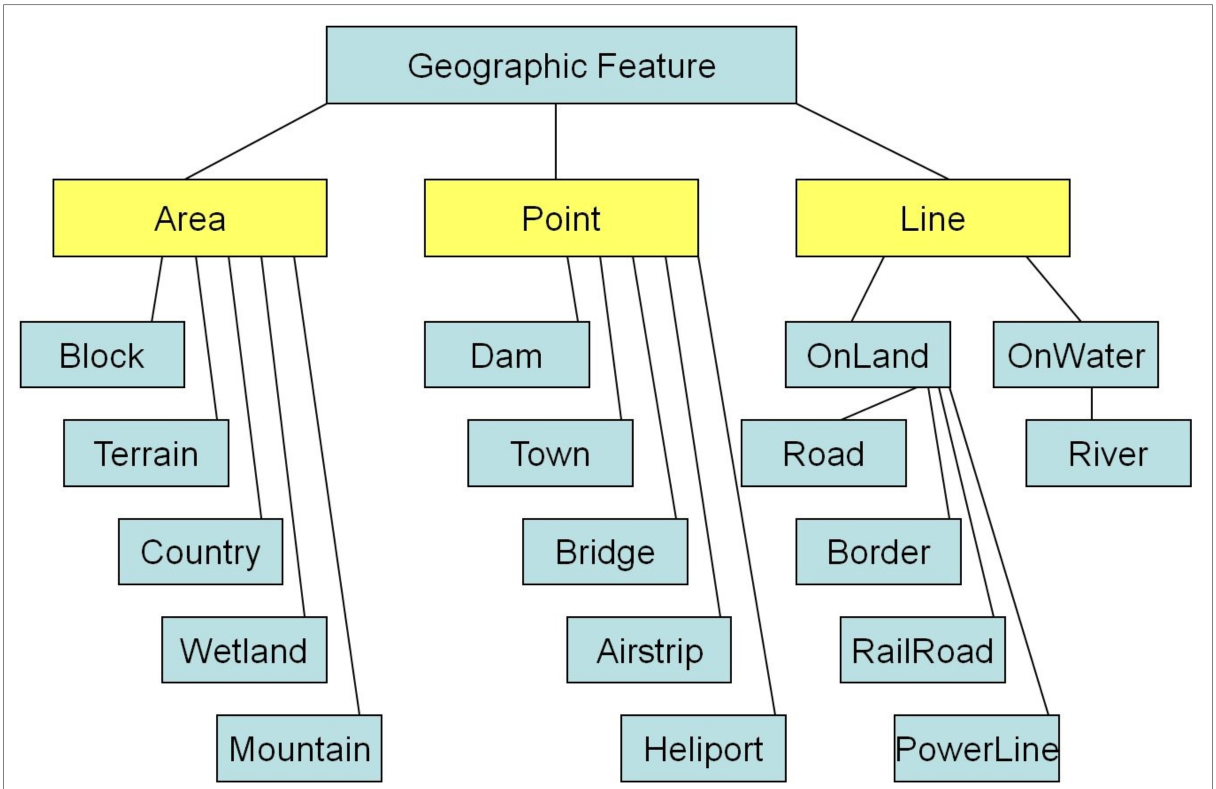


Figure 5. Example of naïve ontologies (Sowa 2005) based on geometric types of features.

3.1 Space Representation and Management

The first question to answer is the status of space since space can intervene in many descriptions. For instance, in the realm of ecology one has to describe biotope, in archeology the importance of places where investigations were made, etc. But in applications such as in environmental and urban planning or in meteorology, space is truly a key-issue. Hence, to put it in a nutshell, we can ask: Is space an attribute of some concepts or rather a new kind of concept?

In various existing ontologies, space can be considered as an attribute with special characteristics, but in other space is really a structuring concept. In geometry, it is common to define 0D (points), 1D (lines), 2D (areas) and 3D (volumes); for an illustration of a geographic ontology based on this principle see Figure 5. But where are points on Earth? The only ones are North and South Pole(s). And what about lines? Mainly imaginary lines such as the equator, meridians and parallels are to be considered. It is true that many geographic information systems (GIS) describe roads and rivers by lines; but in reality, they are areas; the concept of ribbons can be seen as a solution. In addition, some towns can be modeled by points whereas they have also areal extension. Consequently we can ask: Are geometric shapes either a matter of scale or rather have some intrinsic characteristics? The aim of the so-called multi-representation is to offer a solution for this problem.

Furthermore, as some geometric objects have known boundaries such as plots of land, countries, etc., for many geographic features boundaries are indeterminate, for instance mountains, deserts, etc. For these objects, conventional geometry is insufficient to describe them, while fuzzy sets can be an interesting candidate to approach this issues. Figure 6 illustrated the problem mentioned using the example of a river.

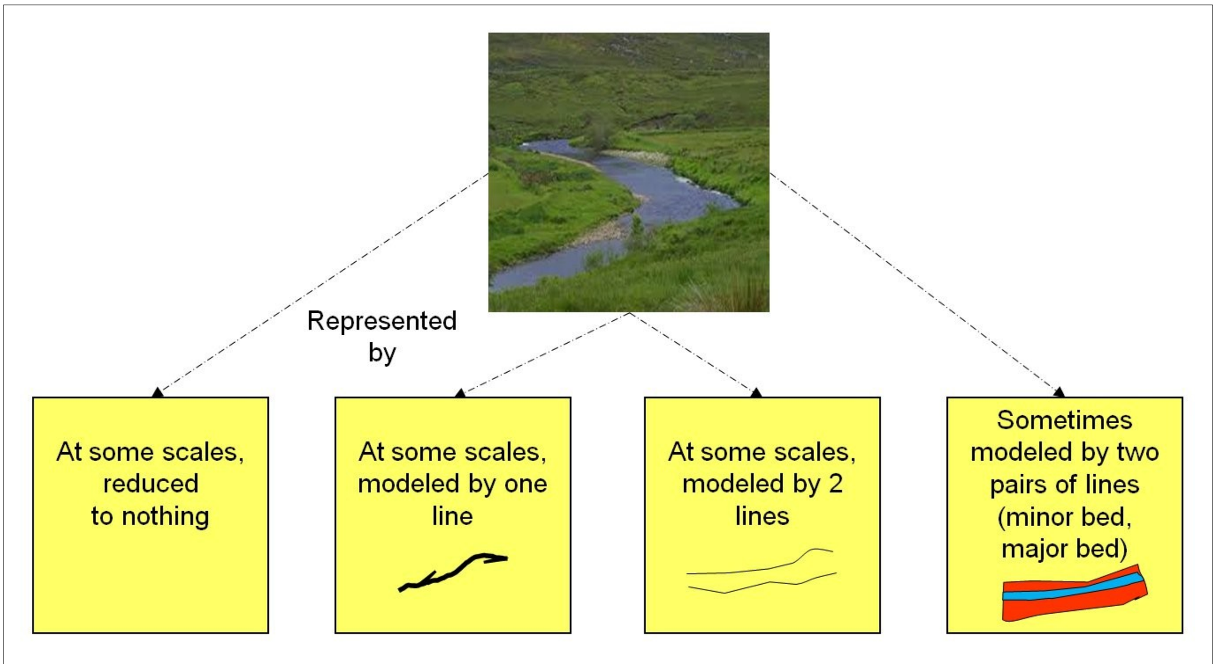


Figure 6. Different geometric representations of a river

3.2 From Spatial to Geographic Relations

As mentioned above, a geographic ontology not only structures geographic objects but also links them by means of spatial relations. A relation type of particular importance are topological relations. Two models can be differentiated: the first one is the so-called

Egenhofer model, shown in Figure 7 (Egenhofer 1994); the second one is known as the RCC model (Randell 1992). Figure 8 illustrates the beginning of a geographic ontology with topological relations.

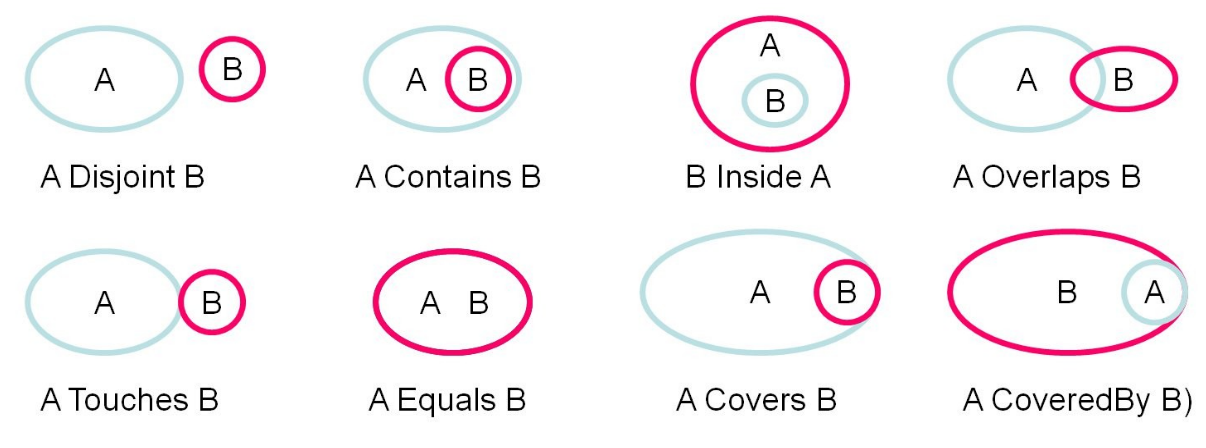


Figure 7. Relations according to Egenhofer (1994).

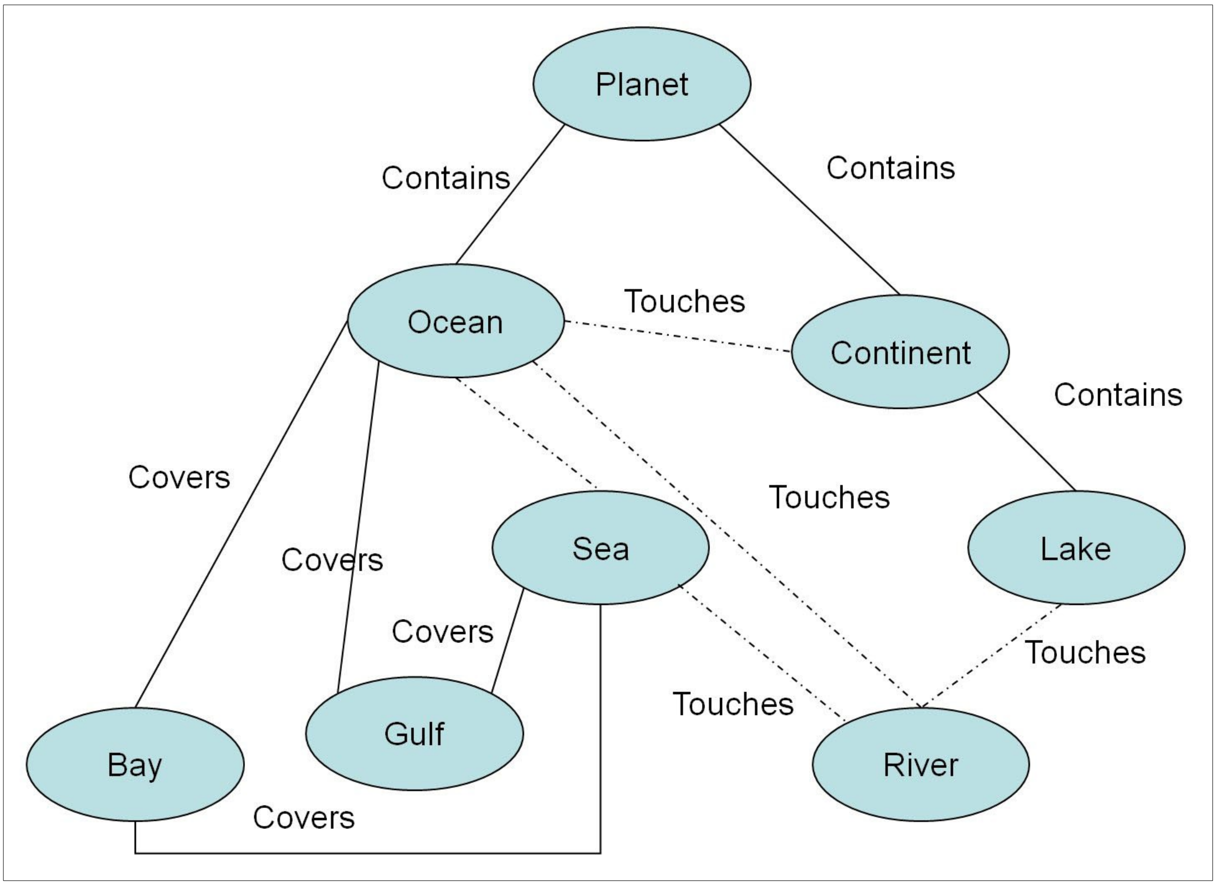


Figure 8. Example of the beginning of a geographic ontology with spatial relations.

3.3 Links with linguistics

Another important issue is language. The scope of the European Towntology project was to design ontologies for urban planning (Teller et al., 2007), focusing its discussion on "either do [...] design a complete ontology in English, and then translate it into various other languages or make several ontologies in different languages and then fusion them into an English ontology?". The problem was not really solved in a satisfactory manner.

One of the problems we have to face is the issue of different languages. E.g., the English term “bank” can represent both a riverside and a financial institution. In other words, the first meaning will be translated in French, and Spanish respectively, by “rive”, “ribera” and the second by “banque” and “banco”. Let us examine another special case: the French term “quai” can refer to a wharf, an embankment, a train platform or a street along a river (see Table 2). In Spanish, especially in Barcelona, “rambla” is a ravine or a special kind of broad avenue. In Venice, “rioterà” is a special type of pedestrian lane whereas other denominations are used such as salizada, sottoportego, ramo, fondamenta, campiello, corte, calle, riva, etc. As far as we know, those terms have no precise equivalent in English.

Table 2: Different meaning of French “quai” - where to integrate it into an ontology?

French	Picture	English	Spanish
		Wharf	Muelle
Quai		Riverside	Avenida a lo largo de un río
		Platform	Andén

As a consequence, different languages can use different concepts to describe features. In others words, two ontologies describing the same domain in different languages can be different. Within international and multilingual projects, this aspect can be difficult to solve.

Furthermore, the same feature can have different names (i.e. exonyms and endonyms) and categories in different languages. Consider e.g. the river Danube: Firstly, in French it is considered not as a river but as a “fleuve” which is defined as a river going to the sea. In other words, there is a topological relation between the river and the sea, notion not integrated into the English term “river”. Moreover, the “Danube” crosses several European countries, taking a different name in almost each country, “Donau” in Germany and Austria, “Dunaj” in Slovakia, “Duna” in Hungary, “Dunav” in Croatia and Serbia, “Dunav” and “Дунав” in Bulgaria, “Dunărea” in Romania and in Moldova, “Dunaj”,and Дунай” in Ukraine. It is also called “Danubio” in Italian and Spanish, “Tonava” in Finnish and “Δούναβης” in Greek, etc. Moreover, its name is feminine in German, while masculine in some other languages.

Gazetteers were initially defined as dictionaries of place names (toponyms), but nowadays an increasing number of databases not only includes feature names but also their types and geometric shapes. Consequently, since each ontology is a knowledge resource organized by concepts and/or types, gazetteers are a knowledge resource based on geographic names.

For instance, the GeoNames[7] database or ontology contains over 10,000,000 geographical names corresponding to over 7,500,000 unique features. All features are categorized into one of nine feature classes and further subcategorized into one of 645 feature codes. Beyond place names in different languages, the database includes latitude, longitude, elevation, population, administrative subdivision and postal codes.

```
<rdf:RDF>
<gn:Feature rdf:about="http://sws.geonames.org/3041565/">
<rdfs:isDefinedBy <rdf:resource="http://sws.geonames.org/3041565/about.rdf"/>
      <gn:name>Andorra</gn:name>
      <gn:alternateName xml:lang="el">Ανδόρα</gn:alternateName>
      <gn:officialName xml:lang="es">Principado de Andorra</gn:officialName>
      <gn:countryCode>AD</gn:countryCode>
      <gn:population>84000</gn:population>
      <wgs84_pos:lat>42.55</wgs84_pos:lat><wgs84_pos:long>1.58333</wgs84_pos:long>
<gn:parentFeature rdf:resource="http://sws.geonames.org/6255148/"/>
<gn:childrenFeatures rdf:resource="http://sws.geonames.org/3041565/contains.rdf"/>
<gn:neighbouringFeatures
rdf:resource="http://sws.geonames.org/3041565/neighbours.rdf"/>
<gn:locationMap rdf:resource="http://www.geonames.org/3041565/principality-of-
andorra.html"/>
</rdf:RDF>

<rdf:RDF>
<gn:Feature rdf:about="http://sws.geonames.org/2510769/">
      <rdfs:isDefinedBy rdf:resource="http://sws.geonames.org/2510769/about.rdf"/>
      <gn:name>Spain</gn:name>
      <gn:neighbour rdf:resource="http://sws.geonames.org/3041565/"/>
      </gn:Feature>
      <gn:Feature rdf:about="http://sws.geonames.org/3017382/">
      <rdfs:isDefinedBy rdf:resource="http://sws.geonames.org/3017382/about.rdf"/>
      <gn:name>France</gn:name>
      <gn:neighbour rdf:resource="http://sws.geonames.org/3041565/"/>
</gn:Feature>
</rdf:RDF>
```

In Laurini (2015), some links between gazetteers, geographic ontologies and multilingualism are studied.

6. Final Remarks and Challenges

To sum up this article, from an IT point of view, the description of any geographic feature must be characterized by the following aspects:

- its names derive from a gazetteer,
- its types and attributes derive from an ontology,
- its relations with other geographic features also derive from an ontology,
- and its geometric description is based on crisp or fuzzy geometries.

One of the key-difficulties identified is the problem of different languages and the cultural aspects behind each language. By using a single language, facets of reality will be forgotten or lost.

[7] http://www.geonames.org/ontology/documentation.html

Several challenges concerning geographic ontologies can be mentioned:

- creating a consensus for the description of geographic features;
- by facing the linguistic problems, deciding whether to create a mono- or multilingual ontology; the solution based on translations from English is not totally satisfying;
- deciding which spatial relations are necessary for an adequate representation of geographic knowledge;
- designing methods for the fusion of existing geographic ontologies, possibly in different languages;
- checking consistency and completeness;

From an application point of view, the challenges could to be as follows:

- What could be the role of geographic ontologies for geographic reasoning, for instance in environmental planning? Moreover: what could be the structure of a geographic inference engine?
- How to use efficiently geographic ontologies for online geographic information retrieval, for instance for tourism purposes?

9. References

Borst, W. N., Akkermans, J. M. & Top, J. L. (1997). Engineering ontologies. *International Journal of Human-Computer Studies, 46*, 365–406.

Cohn, A.G., & Gotts, N.M. (1996). The 'egg-yolk' representation of regions with indeterminate boundaries. In P.A. Burrough & A.U. Frank (eds.), *Geographic Objects with Indeterminate Boundaries* (pp. 171–187), London :Taylor & Francis.

Egenhofer, M. (1994). Deriving the Composition of Binary Topological Relations. J*ournal of Visual Languages and Computing, 5*(2), 133-149.

Fellbaum, C. (ed), (1999). *WordNet: an Electronic Lexical Database.* Cambridge: MIT Press.

Fensel, D. & Groenboom, R. (1997). Specifying Knowledge-Based Systems with Reusable Components. In *Proceedings of the 9th International Conference on Software Engineering and Knowledge Engineering (SEKE-97)*, Madrid, Spain

Fridman, N.& Hafner, C. (1997). The State of the Art in Ontology Design: A Survey and Comparative Review. *AI Magazine, 18*(3), 53-73.

Gruber, T. R. (1993). A Translation Approach to Portable Ontologies. *Knowledge Acquisition, 5*(2), 199-220.

Guarino, N. (1998). Formal Ontology and Information Systems. In N. Guarino (ed.), *Formal Ontology in Information Systems* (pp. 3-15). Amsterdam, Netherlands: IOS Press.

Gómez-Pérez, A. & Corcho, O. (2002). Ontology languages for the semantic web. *IEEE Intelligent Systems, 17*(1), 54-60.

Kavouras, M., Kokla, M., & Tomai, E. (2005). Comparing categories among geographic ontologies, *Computers & Geosciences, 31*(2), 145-154.

Laurini, R. (2014). A Conceptual Framework for Geographic Knowledge Engineering, *Journal of Visual Languages and Computing, 25*(1), 2-19.

Laurini, R. (2015). Geographic Ontologies, Gazetteers and Multilingualism. *Journal of Future Internet, 7*, 1-23.

Lejdel, B., Kazar, O. & Laurini, R. (2015). Mathematical Framework for Topological Relationships between Ribbons and Regions, *Journal of Visual Languages and Computing, 26*, 66-81.

Lord, P. (2010). *Components of an Ontology.* Retrieved December 14, 2015, from Ontogenesis: http://ontogenesis.knowledgeblog.org/514

Randell, D. A., Cui, Z., & Cohn, A. G. (1992). A spatial logic based on regions and connection. *Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning*, San Mateo: Morgan Kaufmann, pp. 165–176.

Smith, M. K., Welty Ch., & McGuinness D. L. (eds.), (2004). *OWL Web Ontology Language Guide*, W3C Recommendation, Retrieved December 14, 2015, http://www.w3.org/TR/2004/REC-owl-guide-20040210/. Latest version available at http://www.w3.org/TR/owl-guide/

Sowa, J. F. (2009). *Building, Sharing, and Merging Ontologies*, Retrieved December 14, 2015 http://www.jfsowa.com/ontology/ontoshar.htm or http://www.jfsowa.com/talks/ontology.htm.

Studer, R,. Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data and Knowledge Engineering (DKE), 25*(1-2), 161-197.

Teller, J., Keita, A., Roussey, C., & Laurini, R. (2007). Urban Ontologies for an improved communication in urban civil engineering projects: Presentation of the COST Urban Civil Engineering Action C21 TOWNTOLOGY. European Journal of Geography *Cybergeo 2007,*(386),

Van Heijst, G., Schreiber, A., & Wielinga, B. (1997). Using explicit ontologies in KBS development. *International Journal of Human-Computer Studies, 46*(2/3), 183–292.

Weibel, R, Keller, S, Reichenbacher, T (1995). Overcoming the knowledge acquisition bottleneck in map generalisation: the role of interactive systems and computational intelligence. In Frank, A U, Kuhn, W. (eds) *Spatial information theory – a theoretical basis for GIS (Proceedings COSIT 95)* (pp. 139–56). Lecture Notes in Computer Science, 988. Berlin: Springer.

Article history:
Received December 14, 2015
Accepted March 3, 2016