Available online at www.sciencedirect.com
ScienceDirect
Electronic Notes in Theoretical Computer Science 168 (2007) 45–59
Electronic Notes in Theoretical Computer Science
www.elsevier.com/locate/entcs

Towards Integration of Geographic Information Systems ⋆

Agustina Buccella[1] Alejandra Cechich[2]

GIISCo Research Group
Departamento de Ciencias de la Computación
Universidad Nacional del Comahue
Neuquen, Argentina

Abstract

Nowadays, geographic information is increasingly used by several entities around the world. Then, the need of sharing information from different sources is an obvious consequence from such proliferation of systems. Unfortunately, integrating geographic information is not a trivial issue. We must deal with several heterogeneity problems, which increase complexity of integration approaches. In order to alleviate some problems, we introduce an integration process based on two main sets of tasks – non-logic and logic. The former is aimed at finding similarities based on structural and syntactic analyzes of geographic data; and the latter is used to calculate inferences from semantics of data by using ontologies. We illustrate the approach by integrating information from two governmental entities, which manage geographic information of North Patagonia, Argentina.

Keywords: geographic information systems, ontology, semantic heterogeneity, description logic

1 Introduction

Currently, newer and better technologies and devices are being created in order to capture a large amount of information about Earth. Nowadays, GPS (Global Positioning System) technology is so common that it is spread all over around such as in cell phones, cars, etc. All of this geographic information is analyzed and stored at different levels of detail in Geographic Information Systems (GIS), possibly distributed on the Web. Then, a fast search for geographic information on the Web will return several links representing different parts of our World. But, what does happen when someone needs information that is divided into more than one system? For example, information about rivers in Patagonia can be obtained by querying

⋆ This work is partially supported by the UNCOMA project 04/E059.

[1] Email: abuccel@uncoma.edu.ar
[2] Email: acechich@uncoma.edu.ar

doi:10.1016/j.entcs.2006.08.023

two or more different systems. Even distribution of information is one of the problems, there are some others: these systems have been developed by different entities with different points of view and vocabularies, and here is when we have to face heterogeneity problems. They are encountered in every communication between interoperating systems, where interoperability refers to interaction between information from different sources involving the task of *data integration* to combine data.

Two systems sharing data representing rivers can be an example to clarify different types of heterogeneity problems as follows [14]: *heterogeneity in the conceptual model* – one system represents a river as an object class and the other as a relationship; *heterogeneity in the spatial model* – rivers can be represented by polygons (or a segment of pixels) in one system, while they are represented by lines in the second system; *structure or schema heterogeneity* – both systems hold the name of a river but one keeps information about the border; and *semantic heterogeneity* – one system may consider a river as a natural stream of water larger than a creek with border and the other defines a river as any natural stream of water reaching from the sea, a lake, etc. into the land.

In this work we focus on the last problem, semantic heterogeneity, during the schema integration of different but related information sources. Generally speaking, two essential tasks are involved in the semantic integration process: *semantic enrichment* and *mapping discovery* [34]. The main goal of *Semantic Enrichment* is to reconcile semantic heterogeneity, so it involves adding more semantic information about the data. Various proposed approaches add extra semantic information through the use of metadata or ontologies. We are particularly interested in those using ontologies because, by definition, they provide a vocabulary to represent and communicate knowledge about the domain and a set of relationships containing the terms of the vocabulary at a conceptual level. Ontologies are actually extensively proposed as tools to face heterogeneity problems. For example, different proposals are using formal ontologies to enrich the conceptual schema and thus to improve the integration process [7,8,14,15].

The semantic enrichment task is essential to reach the second task, *Mapping Discovery*. Several surveys [16,18] have been presented analyzing different proposals related to semantic matching, i.e. building of mappings. As we build our process upon finding of mappings, we further describe some of these proposals in Section 2.

Our approach is focused on searching mappings between two geographic ontologies. These mappings are used to generate a global integrated ontology containing the structure of data of a federated system, which includes a set of distributed and autonomous sources of data.

Our proposal is at schema level, that is, instances of concepts are not taken into account but only the structure of data. From these data, formal ontologies are built based on geographic standards and using classes, properties, constraints, etc. All of these elements of the ontologies are taken into account in the process of searching mappings. Novelty of our proposal bases on that not only taxonomic relationships are considered but also more complex elements of the ontologies.

Additionally, we should consider the following hypothesis tested by Mark et al.: *"geographic and non-geographic entities are ontologically distinct in a number of ways"*. Their experiments tested the degree to which ordinary people can code the geographic domain at conceptual level. As a conclusion of this study, there is a set of geographic terms that have higher frequency, that is, they are more recurrent terms. In principle, by knowing which these terms are, the similarity process could be simplified.

As formal ontologies are used by our proposal, a logical formalism is needed to represent them. As Cocchiarella wrote [3] *"Formal ontology is a discipline in which the formal methods of mathematical logic are combined with the intuitive, philosophical analyzes and principles of ontology"*. In particular, Description Logic [2] has been selected to formalize our ontologies. Besides capabilities any logic language provides, reasoning systems capable of processing such formalism exists. Thus, our proposal takes advantage of these features and applies them in the inferred mapping process.

This paper is organized as follows: next Section presents related works in the literature together with relations to our approach. Then, an overview of the architecture briefly describing the main components of our proposal is presented in Section 3, followed by a description of our integration process. Then, a case study illustrating the application of our method is shown in Section 4. Future work and conclusions are discussed afterwards.

2 Related Work

Mapping discovery by using ontologies has being extensively investigated during the last years. Various approaches have emerged proposing processes and techniques to find similarities between elements of different but related ontologies. Some approaches [4,21,24,25,10] involve non-specific information systems and others [14,30,31,34,36] are specifically oriented towards geographic information. Both groups complement each other because solutions proposed by some of them can be used by others. For example, the similarity functions proposed in [31] have been applied in [4].

With respect to the first group, OBSERVER [25] is one of the approaches more referenced in the literature. It defines a model for dealing with multiple ontologies avoiding problems about integrating global ontologies. One important component in OBSERVER is the IRM (Interontology Relationships Manager) shared repository. It can be seen as a catalog of semantics of the system used to solve the "vocabulary problem" (heterogeneous vocabularies used to describe the same information). The IRM component supports ontology-based interoperation by defining several kinds of interoperable relationships as synonym, hyponym, hypernym, overlap, etc. among

[3] http://www.stoqnet.org/lat_notes.html

the terms of different (locally developed) ontologies. These relationships are necessary to map the elements of the ontologies, but OBSERVER does not propose a semantic matching method, so the mappings must be found manually.

Additionally, examples of semantic matching methods are described in [4,21,24,10]. For example, [10] and [24] propose two similar ontology-merging tools. On one hand, the PROMPT tool described in [10] proposes an interactive tool that guides the user through the merging process. However the main problem with the PROMPT tool is that it is highly dependent on the names of the concepts in the ontology. On the other hand, Chimarea [24] provides support for merging of ontological terms from different sources, checking the coverage and correctness of ontologies and maintaining ontologies over time. Except for several situations referring to structural aspects of the ontologies, Chimarea does not make any suggestion to the user; and the only relation that Chimarea considers is the subclass/superclass relation.

Another proposal for semantic matching is introduced in [21], where a lexical and a conceptual layer are used to find similarities. At the lexical level, the method uses a lexical function called lexical similarity measure (SM). At the conceptual level, concepts (classes and properties) are compared taking into account the taxonomies in which they appear. However, some types of properties are not considered by this method.

With respect to the second group, which involves geographic information, the work in [30] presents a combination of two different approaches to similarity assessment – the feature matching process [35] and the semantic distance. Common features increase the similarity value and distinct features decrease it. The main disadvantage with this method is that the similarity values cannot be calculated neither automatically nor semi-automatically due to the high dependence on natural language descriptions.

Another example is the work introduced in [14],which defines a method by using ontologies represented by a logical language to integrate data base schemata. A reasoning system is used to merge ontologies based on a set of inferred similarity relations. Source ontologies are built based on the elements of source schemas and taking into account a higher-level ontology (which is an ontology with general terms and minimum constraints). The results are formal ontologies written in description logic. As these source ontologies are based on a higher-level ontology (that is, they include this ontology in their definitions and add more detailed definitions for some other elements), the reasoning system only has to find similarity relations between them. The PowerLoom [20] reasoning system was used to evaluate DL definitions.

In the ODGIS (Ontology-Driven Geographic Information Systems) approach [6] authors introduce a framework for the integration of geographic information. This framework has two main aspects, the *knowledge generation* in which the ontologies are specified and the *knowledge use* in which a group of components interact to answer a query (by using mechanisms to retrieve instances of instances of classes from ontologies). Particularly in the first one, ontologies of different level of detail are specified – a high-level ontology (at the top level), domain ontologies (based on the the previous one), and task ontologies are some of them. The main contribution of this approach is the use of roles denoting different functions an object can take depending on the perspective. Thus, each entity of an ontology can play many roles. Roles and hierarchy mechanisms are used as a tool to integrate the different ontologies.

Finally, in [34] by using conceptual models and description logics, a methodology for the integration of spatio-temporal conceptual schemas is defined. The sources to be integrated are represented by using the MADS conceptual data model (Modeling of Application Data with Spatio-temporal features [27]) with its multiple representation capabilities [26] in order to manipulate geographical information through multiple perspectives of the same information. The proposal is based on the specification and use of inter-schema knowledge, that is, they are not focused on the semantic matching activity but on a methodology to build an integrated system. An expert designer is responsible of finding the possible mappings between two MADS conceptual models. Description logic reasoning services are used to check the satisfiability of the set of inter-schema mappings. The model has a rich spatio-temporal semantics, but the mappings must be discovered manually.

By implementing two kind of processes, logic and non-logic, our method combines two aspects of the integration problem. Firstly, a formal language is used in order to represent the ontologies. These "formal ontologies" are built based on the standards defined for geographic information systems. By representing ontologies with a formal language, we find new relations taking advantage of the logic of data. A reasoning system is used to perform inferences about the defined taxonomy. Thus, implicit relations are found and added to the ontologies.

The second important aspect is the syntactic and structural analysis of the elements of the ontologies. These analyzes are aimed at finding similarities that can not be found by the logic process. Therefore, both processes complement each other.

Comparing our proposal with those in the literature, some points in common can be found with respect to the proposal described in [14]. For example, Description Logics is used as a tool to represent domain ontologies in order to take advantage of the inference mechanisms. But in [14], domain ontologies must commit to the same high-level ontology to allow the reasoning system to start the integration process. Thus, domain ontologies are not independent because different communities must agree with the high-level ontology. Another proposal with some similarities with respect to ours is the ODGIS approach [6]. The way semantic granularity is managed by this approach is applied to our methodology (looking for the first possible intersection going upward in the ontology trees); however we do not use ontologies of different levels of detail. As differences, this proposal needs human intervention (on the ontology creation and on the role definitions) and uses a reduced set of semantic relations (is-a, part-of, and whole-of). Besides, as in [14], domain ontologies are based on a high-level ontology.

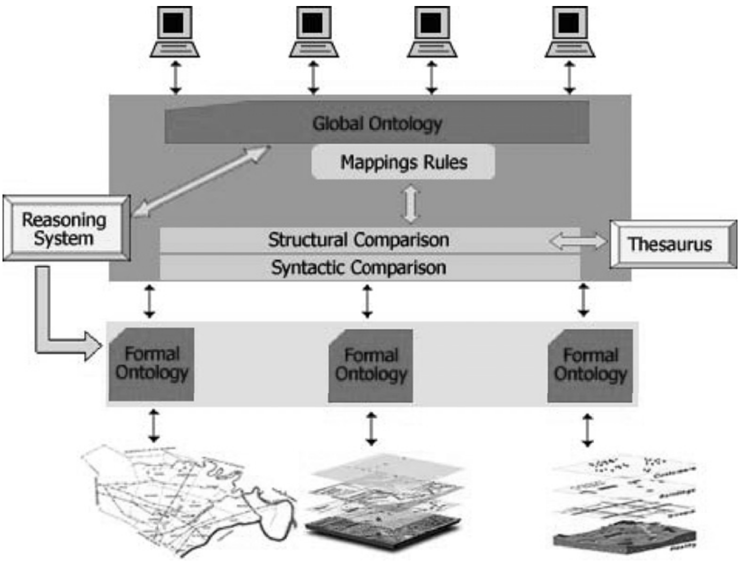

Fig. 1. Architecture of our Proposal

3 Overview of our Architecture

In previous works [4,5], we have proposed a layered-based architecture (Figure 1) to integrate different information sources by using ontologies. Besides, an ontology-merging method based on the components of this architecture has been defined. There, the ontologies can be non-formal because syntactic and semantic relationships are only taken into account. The method contains three levels (syntactic, semantic and user level) allowing a user to find several correct mappings.

Now, to instantiate our approach to deal with geographic information, source information is represented by local and autonomous geographic information systems as parts of the first layer. Formal ontologies are in the next layer representing information extracted from each source. Logic is used as a formalism for ontology representation by using for example Description Logic (DL) [2] and Frame-based Logic (FLogic) [17]. An exhaustive comparison between them can be found in [14]. Due to several advantages of DL with respect to FLogic, DL has been selected to represent our ontologies. Besides, a reasoning system has been used in order to perform inferences about semantics of data.

To build the integrated system, the next layer defines six main components. As a result, a global ontology is built involving the concepts included in the formal ontologies. In order to query the system, potential users browse this global ontology.

Let us briefly describe the components' profile. The *Reasoning System* and *Thesaurus* are external components because they are out of the scope of our development, although they are part of the integration process indeed. The Reasoning System is used to perform inferences between ontologies such as inferred subsumption relations between concepts. Thesaurus is used by the non-logic process in which the mappings are found by comparing concepts syntactic and structurally. Our integration process is divided into two main sets of tasks – *logic* and *non-logic*.

Logic tasks involve the *Reasoning System* aforementioned. Non-logic tasks involve the *Syntactic Comparison*, *Structural Comparison* and *Thesaurus* components. The Syntactic Comparison component analyzes concepts from different ontologies syntactically. That is, the name of concepts are compared. Following, the Structural Information component analyzes relationships and attributes of the elements of the ontologies. Besides, the Thesaurus component is used to extract semantic information in order to find synonym relationships. By using the syntactic and thesaurus results, and performing a structural matching, the Structural Comparison component generates a set of mappings relating source concepts.

Next Section contains a detailed explanation about our integration process.

3.1 The Integration Process

The geographic information is the base of our system. This information is represented by using formal ontologies in order to take advantage of the semantic of data. By using geographic standards such as ISO/IS 19107 [12], ISO/DIS 19109 [11], etc., formal ontologies are created. Firstly, a conceptual model is built capturing the information represented by the geographic sources. And later, these conceptual models are translated to formal ontologies by using any ontology editor (in this case, Protégé [28]). Thus, these ontologies are then inputs of our integration process.

The integration process analyzes them taking into account classes, properties (associations and attributes), specialization/generalization relations and constraints of the ontologies [4].

Two main sets of tasks are involved in this process – *non-logic* and *logic* tasks – each of them focusing on different mismatches between the two ontologies. The first group, *non-logic tasks*, takes into account *syntactic* and *structural* analyses involving five components: Syntactic Comparison, Structural Comparison, Thesaurus, Mappings Rules and Global Ontology (Figure 1).

In the *syntactic analysis* (performed by the Syntactic Comparison component) two formal ontologies are analyzed looking for syntactic similarities between elements involved in them. Three similarity functions are used here as follows.

- The *edit distance* function, which considers the number of changes that must be done to turn one string into the other, and weights the number of these changes with respect to the length of the shortest string.
- The *trigram* function [19], which is based on the number of different trigrams in two concepts or strings.
- And the *check constrains* function, which compares the constrains applied to the properties, for example, cardinality constrains. Only when both properties have the same restrictions, the function returns 1; otherwise it returns a percentage according to the number of restrictions that are the same.

The first two functions compare the names of the concepts in a different way. Thus, both functions return a different similarity result depending on the syntaxis of the compared names.

Following, in the *structural analysis* (performed by the Structural Comparison component) two main steps are carried out. In the first step, the results of the syntactic analysis are combined together with the results from the thesaurus information. The Thesaurus component is used to extract synonym relationships between the concepts of the ontologies. These relationships are necessary because synonyms (in general) are not similar syntactically. In this case, EuroWordNet [4] is used as the thesaurus. EuroWordNet is a multilingual database with wordnets for several European languages, such as Spanish, that is the one we are interested in. It is based on the American wordnet for English (WordNet [29]). However, the main problem we have to face at this point is that not every wordnet in WordNet is in EuroWordNet; only around 35% of the wordnets is translated. Thus, there are several words representing elements in the formal ontologies without synonym relations. Besides, EuroWordNet (like WordNet) is not specifically for geographic information and when some specific geographic words are used, they cannot be found. Nevertheless, to the best of our knowledge geographic thesauruses have not yet been created.

In the second step, in order to perform an structural comparison, the similarity function described in [3,31] is used. This function compares the number of properties that the classes have in common and analyzes them in a hierarchy (by calculating the depth of the most common superclass between the classes). For example, if a "Country" class is described by three attributes (name, border and inhabitants) in one ontology, and in the other the same class is also described by three attributes (name, border and isCapital), the function returns $\frac{2}{3}$ (when these classes are in the same depth in the hierarchy).

Thus, with the results of the last analysis (structural analysis), mapping rules are generated. They are used to build the global ontology which is browsed by users in order to query the integrated system.

Now, we should look at the logic tasks. Three components of our architecture are involved in the logic process (Figure 1): Reasoning System, Formal Ontologies and Global Ontology. This process takes place in two different moments within the whole integration process, *before* and *after* the non-logic tasks. In both moments, the Reasoning System component is used to take advantage of the logic of data. Remember that we are working with formal ontologies which are represented by a logical language (in this case, Description Logic). Therefore, by means of a reasoning system, such as RACER [13], inferences over the ontologies can be performed. We take advantage of the capability of inferring subsumption relations between classes and properties in the schema (TBox). That is, the reasoning system will determine where a concept can be located in a taxonomy hierarchy (a hierarchy built by means of a subconcept relation).

Before the non-logic tasks, the logic process is performed to recognize subsumption relations in each formal ontology. That is, the reasoning system checks for class subsumption on the formal ontologies. Moreover, equivalent classes can be found within an ontology because two concepts C and D are equivalent when C is

[4] http://www.illc.uva.nl/EuroWordNet/

subsumed by D and D is subsumed by C [2]. Thus, each formal ontology enters in the non-logic process with an inferred hierarchy (computed by the reasoner).

After the non-logic tasks, the reasoning system is used to analyze possible subsumption relations in the global ontology. Remember that this ontology has been generated by using the mapping rules obtained by the non-logic process. Thus, the global ontology will also contain the inferred hierarchy.

Besides, the reasoning system is used to check the consistency of each model, the formal ontologies, and the resultant global ontology. Here, the validity of intentional definitions (in TBOX) is checked. If an inconsistency is found, an expert user is responsible of solving it.

4 A Case Study

In order to illustrate the process, two different geographic information systems have been selected to participate in the integration. These systems are not currently working together but they store similar information. One system is managed by the AIC (Autoridad Interjurisdiccional de Cuencas) entity which is in charge of managing, controlling, using and preserving the basins of the rivers Neuquén, Limay and Negro in Patagonia, Argentine. The covered area includes the Río Negro, Neuquén and part of Buenos Aires provinces (about 140.000 Km^2 representing the 5 % of the Argentinean total territory). The another system is implemented by the Provincial Office of Territorial Cadaster (DPCT - Dirección Provincial de Catastro Territorial) in which all information about buildings, streets, parcels, etc. is stored. The covered area includes only the Neuquén Province (about 94.068 Km^2).

Figure 2 graphically shows a simplified part of the conceptual model of the AIC system. The map shows some of the real objects entities represented (dot arrows references them). Three geometric types are used in this model and references are shown in the right side of the figure.

Figure 3 graphically shows part of the other conceptual model, the DPCT system. Both models have different granularity because of their different interests. Granularity here refers to semantic granularity including the level of detail involved in the selection of features [9]. The AIC system represents natural objects (rivers, lakes, mountains, etc.) with more detail than the DPCT system. On the other hand, objects such as buildings, owners and city organization are more detailed in the DPCT system.

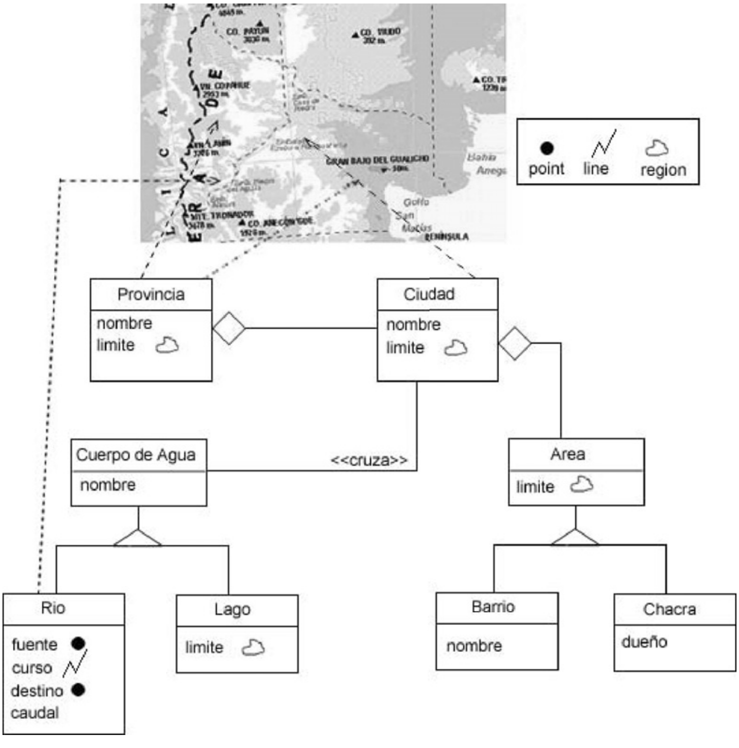

Fig. 2. Conceptual Model for AIC system

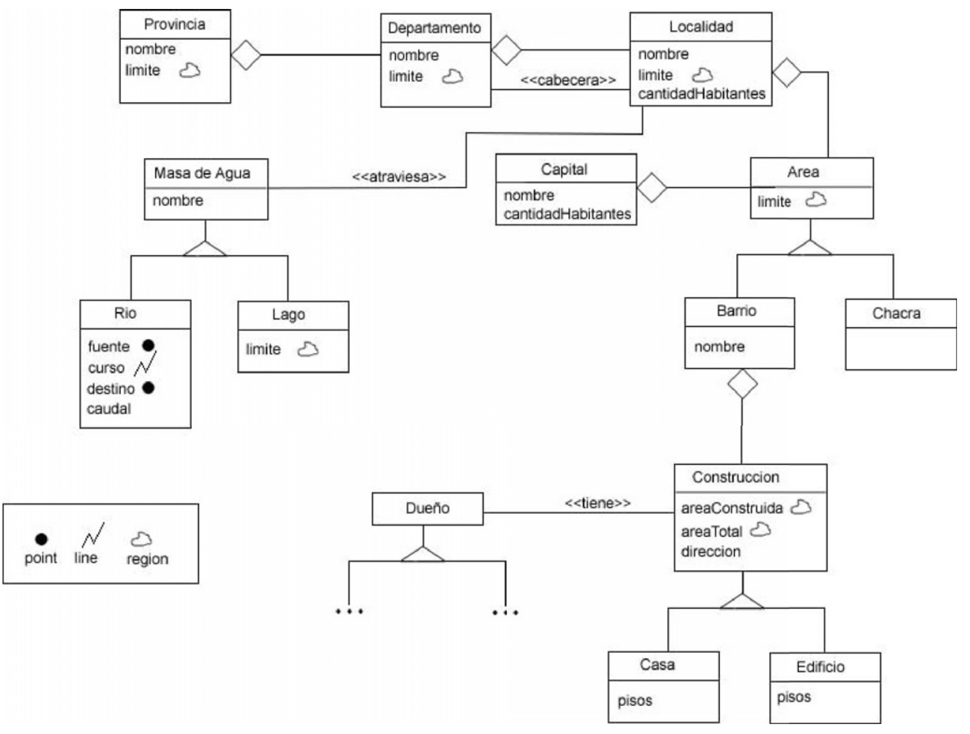

Fig. 3. Conceptual Model for DPCT system

Granularity is managed by using generalization/specialization relationships [9]. When the information about these relationships is not enough, EuroWordNet is used as a tool to obtain them. Besides, relations of hypernyms and hyponyms are browsed in this situation.

The *logic process* is the beginning of the integration process between these two ontologies. We represent both conceptual models by using an ontological language (in this case, OWL [1] and Protégé [28] to model the ontologies). The geographic data types have been imported from the iso-19107 [5] ontology. They are part of the feature geometry model [32] used to represent geographical types.

For example, in the case of the DPCT conceptual model (Figure 3) the "Capital"(representing Capitals of a province or state) and "Localidad" (representing cities of a province) classes are defined as following:

```
<owl:Class rdf:ID="Localidad">
     <owl:equivalentClass>
        <owl:Restriction>
           <owl:onProperty rdf:resource="#compuestoPor"/>
           <owl:allValuesFrom rdf:resource="#Area"/>
        </owl:Restriction>
    </owl:equivalentClass>
    <rdfs:comment rdf:datatype="&xsd;string"
    >Parte en que se divide un territorio</rdfs:comment>
</owl:Class>

<owl:Class rdf:ID="Capital">
   <rdfs:subClassOf rdf:resource="&owl;Thing"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#compuestoPor"/>
        <owl:allValuesFrom rdf:resource="#Area"/>
      </owl:Restriction>
    </rdfs:subClassOf>
  <rdfs:comment rdf:datatype="&xsd;string"
    >Poblacin principal de un pas o de una provincia</rdfs:comment>
</owl:Class>
```

The "Area" class represents different areas that compose a city or a capital of a province (represented by the "compuestoPor" property). These two definitions of the classes generate an inferred hierarchy in which "Capital" is proposed as subclass of "Localidad". "Localidad" is a *defined class* containing one necessary and sufficient restriction (the "compuestoPor" property). Thus, any individual that satisfies the definition will belong to the class. Otherwise, this definition is a necessary condition for "Capital". Note that if the "compuestoPor" property is a necessary condition in both classes, the reasoning system computes an equivalence relation.

Once the classification of the taxonomy is applied separately to both formal ontologies, the *non-logic process* starts. To do so, both ontologies are analyzed syntactically by comparing the names of the elements. Three functions are used to find similarities between them. For example, when "Ciudad" (representing cities) of the first ontology (AIC ontology, Figure 2) and "Localidad" of the other (DPCT ontology, Figure 3) are compared, the results of the two first syntactic functions are very low because these concepts are not similar syntactically. The edit distance function returns 0, and the trigram function returns $\frac{1}{10}$. Finally, the check constrain

[5] http://loki.cae.drexel.edu/ wbs/ontology/iso-19107.htm

function is equal to 1 because the constrains applied to these classes are the same.

Following, during the structural analysis, synonym relations are extracted from the thesaurus to determine if the compared elements are related. In this example, these classes contain a synonym relation.

The second step of the structural analysis compares only the classes calculating the number of properties they have in common. Following our example, the "Ciudad" and "Localidad" classes are defined with "nombre" (denoting the name of the city), "limite" (denoting the boundary of the city), and "compuestoPor" (denoting the areas that compose a city) as properties. In the case of the DPCT ontology, the "Localidad" class is a superclass of the "Capital" class. Besides, as these classes are in the same level in the hierarchy, a very high result is returned by the structural function.

Then, all the information of both analyzes (syntactic and structural) are combined in order to obtain similarity values as results. More specifically, these results are parts of similarity functions in which a sum of products ($value \times weight(w)$) is performed. In this case study we consider that the weights (w values) in the similarity functions are evenly distributed (the sum of weights is equal to 1). But these weights can change accordingly with the importance that an expert user wants to give them. Then, the mapping rules are generated by using the results of these similarity functions.

Finally, when the mapping rules have been generated and an initial global ontology has been built, the logic process starts again in order to perform classification tasks. In this step and as before, subsume relations can be found including subclass/superclass and equivalence relations. Besides, if the global ontology is found to be inconsistent (due to the result of some mapping rule), an expert user is needed to solve the problem. So, he/she has to reconsider some decisions of the non-logic analysis.

Table 1 shows the mapping rules found for the classes in the case study presented. The properties "cruza" (meaning crosses) and "atraviesa" (meaning traverses) are found similar because they are synonyms. Then, all the equally named properties are also found as similar.

$Provincia \doteq Provincia$	$Chacra \doteq Chacra$
$Cuidad \doteq Localidad$	$Lago \doteq Lago$
$CuerpoDeAgua \doteq MasaDeAgua$	$Rio \doteq Rio$
$Area \doteq Area$	$Barrio \doteq Barrio$

Table 1
Mappings Rules generated by our Integration Process

By applying our method in this case study, we can see the set of mappings that are possible to be found taking into account all the elements of the ontologies. Ontologies are not only taxonomies but also properties denoting more semantics. These properties (used in all conceptual models) are compared in the non-logic process.

Moreover, a logic process is performed in order to infer implicit relations that can not be found by the non-logic process. Therefore, by combining the two processes our method allows to discover more suitable mappings than similar approaches, such as the ones cited in Section 2.

5 Conclusion and Future Work

In this work an architecture and a process to integrate geographic sources have been described. In particular, we work with two geographic sources managed by two organizations of our region. Our method is aimed to assist part of the whole integration process giving solutions to the construction of the new system. The main advantage of our method is the combination of two processes, logic and non-logic, which complement each other. That is, mappings or implicit relations found in one process are taken into account by the other. In this way, more properties of the ontologies are represented and compared.

As a future work we are working on two processes. Firstly, on the logic part, inconsistencies are being analyzed in order to investigate automatic options to solve them. Remember that currently they are manually solved by an expert user. On the other hand, on the non-logic part, NLP (Natural Language Processing) techniques are being considered as tools to find one-to-many mappings; that is, one element of one ontology might be mapped to two elements of the other. A preliminary work by using NLP tools can be found in [33] where the "multiconcept" definition (MCR)[22], is used in order to find the *complex mappings*. Besides, we are working on defining a set of structural terms frequently used in geographic domains. It is a similar study to the one introduced in [23], but using the Spanish language instead.

References

[1] Antoniou, G. and F. van Harmelen, *Web Ontology Language: OWL*, in: S. Staab and R. Studer, editors, *Handbook on Ontologies in Information Systems* (2003).

[2] Baader, F., D. Calvanese, D. McGuiness, D. Nardi and P. Patel-Schneider, editors, "The Description Logic Handbook – Theory, Implementation and Applications," Cambridge University Press, 2003.

[3] Buccella, A., A. Cechich and N. R. Brisaboa, *A Context-Based Ontology Approach to Solve Explanation Mismatches*, in: *JCC'03 Jornadas Chilenas de Computación*, Chillan, Chile, 2003.

[4] Buccella, A., A. Cechich and N. R. Brisaboa, *A federated layer to integrate heterogeneous knowledge*, in: F. Gadducci and M. H. ter Beek, editors, *VODCA'04: First International Workshop on Views on Designing Complex Architectures*, number 142 in ENTCS, Elsevier Science B.V, Bertinoro, Italy, 2004, pp. 79–97.

[5] Buccella, A., A. Cechich and N. R. Brisaboa, *A Three-Level Approach to Ontology Merging*, in: *MICAI'05: Fourth Mexican International Conference on Artificial Intelligence*, LNCS **3789** (2005), pp. 80–89.

[6] Fonseca, F., "Ontology-driven Geographic Information Systems," Ph.D. thesis, University of Maine (2001).

[7] Fonseca, F., C. Davis and C. Camara, *Bridging ontologies and conceptual schema in geographical information integration*, Geoinformatica (2003), pp. 307–321.

[8] Fonseca, F., M. Egenhofer, P. Agouris and C. Camara, *Using ontologies for integrated geographic information systems*, Transactions in GIS **3** (2002).

[9] Fonseca, F., M. Egenhofer, C. Davis and C. Camara, *Semantic Granularity in Ontology-Driven Geographic Information Systems*, Annals of Mathematics and Artificial Intelligence **36** (2002), pp. 121–151.

[10] Fridman Noy, N. and M. A. Musen, *PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment*, in: *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence* (2000), pp. 450–455.

[11] Geographic Information, Rules for Application Schema. Draft International Standard 19109, ISO/IEC (2002).

[12] Geographic Information, Spatial Schema. International standard 19107, ISO/IEC (2003).

[13] Haarslev, V. and R. Moller, *RACER System Description*, in: P. Lambrix, A. Borgida, M. Lenzerini, R. Moller and P. Patel-Schneider, editors, *Proceedings of the CEUR-WS International Workshop on Description Logics*, 22, Linkoeping, Sweden, 1999.

[14] Hakimpour, F., "Using Ontologies to Resolve Semantic Heterogeneity for Integrating Spatial Database Schemata," Ph.D. thesis, Zurich University (2003).

[15] Hakimpour, F. and A. Geppert, *Global Schema Generation using formal ontologies*, in: *Proceedings of ER'02*, LNCS **2503**, 2002, pp. 307–321.

[16] Kalfoglou, Y. and M. Schorlemmer, *Ontology mapping: the state of the art*, The Knowledge Engineering Review **18** (2003), pp. 1–31.

[17] Kifer, M., G. Lausen and J. Wu, *Logical Foundations of Object-Oriented and Frame-Based Languages*, J. ACM **42** (1995), pp. 741–843.

[18] Klein, M., *Combining and relating ontologies: an analysis of problems and solutions*, in: *Proceedings of the IJCAI-2001 Workshop on Ontologies and Information Sharing*, Seattle, USA, 2001.

[19] Lin, D., *An Information-Theoretic Definition of Similarity*, in: *Proceedings of the Fifteenth International Conference on Machine Learning*, 1998, pp. 296–304.

[20] MacGregor, R., H. Chalupsky and E. Melz, *PowerLoom Manual*, Technical report, University of Southern California, http://www.isi.edu/isd/LOOM/Power-Loom/documentation/manual.pdf (1997).

[21] Maedche, A. and S. Staab, *Measuring Similarity between Ontologies*, in: *Proceedings of EKAW'02: the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*, LNCS (2002), pp. 251–263.

[22] Magnini, B., M. Speranza and G. Girardi, *A semantic-based approach to interoperability of classification. Hierarchies: Evaluation of Linguistic Techniques*, in: *Proceedings of COLING'04*, Geneva, Switzerland, 2004.

[23] Mark, D. M., A. Skupin and B. Smith, *Features, Objects, and Other Things: Ontological Distinctions in the Geographic Domain*, in: *Spatial Information Theory: Foundations of Geographic Information Science*, LNCS **2205** (2001), pp. 488–502.

[24] McGuinness, D., R. Fikes, J. Rice and S. Wilder, *An environment for merging and testing large ontologies*, in: *Proceedings of KR'00*, 2000, pp. 483–493.

[25] Mena, E., A. Illarramendi, V. Kashyap and A. P. Sheth, *OBSERVER: An Approach for Query Processing in Global Information Systems Based on Interoperation Across Pre-Existing Ontologies*, Distributed and Parallel Databases **8** (2000), pp. 223–271.

[26] Parent, C., S. Spaccapietra and E. Zimányi, *The murmur project: Modeling and querying multi-representation spatio-temporal databases*, Information Systems (2005).

[27] Parent, C., S. Spaccapietra and E. Zimányi, *Spatio-temporal conceptual models: data structures + space + time*, in: *Proceedings of GIS'99: the 7th ACM international symposium on Advances in geographic information systems* (1999), pp. 26–33.

[28] Protégé, http://protege.stanford.edu/doc/users_guide/index.html (2000).

[29] Richardson, R. and A. Smeaton, *Using WordNet in a Knowledge-Based Approach to Information Retrieval*, Technical Report CA-0395, Dublin City University, School of Computer Applications, Dublin, Ireland (1995).

[30] Rodriguez, A. and M. Egenhofer, *Comparing Geospatial Entity Classes: An Asymmetric and Context-Dependent Similarity Measure*, International Journal of Geographical Information Science **18** (2004), pp. 229–256.

[31] Rodríguez, A. and M. J. Egenhofer, *Determining Semantic Similarity among Entity Classes from Different Ontologies*, IEEE Transactions on Knowledge and Data Engineering **15** (2003), pp. 442–456.

[32] Rodríguez Luaces, M. A., "A Generic Architecture for Geographic Information Systems," Ph.D. thesis, Univerdade da Coruña (2004).

[33] Roger, S., A. Buccella, A. Cechich and M. Palomar, *ASeMatch: A Semantic Matching Method*, in: *TSD'06: Ninth International Conference on Text, Speech and Dialogue*, to appear, Brno, Czech Republic, 2006.

[34] Sotnykova, A., C. Vangenot, N. Cullot, N. Bennacer and M. Aufaure, *Semantic mappings in description logics for spatio-temporal database schema integration*, Journal on Data Semantics III (2005), pp. 143–167.

[35] Tversky, A., *Features of Similarity*, Psychological Review **84** (1977), pp. 327–352.

[36] Visser, U., "Intelligent Information Integration for the Semantic Web," LNCS **3159**, Springer Berlin - Heidelberg, 2004.