Syntactic and Semantic GDB Conceptual Schemas Integration

Guillermo Nudelman Hess, Cirano Iochpe
Universidade Federal do Rio Grande do Sul – Instituto de Informática
Porto Alegre, Brazil
{hess,ciochpe}@inf.ufrgs.br

SUMMARY

This paper presents an architecture and a methodology to integrate GDB conceptual schemas with the goal to enable the reuse as well as the interchange of them or part of them. The integration is separated in two levels. The syntactic integration consists on transforming the conceptual schemas representation to a canonical format, the GML 3.0. After that the semantic integration is performed, presenting an algorithm to solve the semantic heterogeneities between the components of the conceptual schemas. This algorithm is based on an ontology and on the application of some similarity matching formulas.

KEYWORDS: *GDB schema integration, Ontology, Similarity matching*

INTRODUCTION

The number of organizations and departments adopting Geograhic Information Systems (GIS) testifies the popularization of the GIS. Because of the multidisciplinary characteristic of the GIS, the community's growth leads to a scenario where many times users of these systems are from areas other than computer science, such as geography, geology, urban planning, etc. Therefore the GIS systems should provide means to automate some tasks which are beyond the users knowledge and compose the background of the applications. One of these tasks is the geographic database (GDB) design.

The GDB conceptual modeling is quite more complex than the traditional (descriptive) database's design because of the spatio-temporal data. On the other hand, since the geographic reality is quite stactic in terms of phenomena and components, the databases schemas tend to be quite similar, what leads to a scenario where at least parts of schemas may be reused (Burrough, McDonnell, 1997) in the modeling process.

These two characteristics of the GDB design associated to the fact that the one who is using the GDB may not be familiar with the GDB design motivated us to develop an architecture and a methodology to integrate GDB conceptual schemas. Furthermore we make them available to reuse by offering a single canonical format to describe their structure and a unification of language to describe their content (the represented phenomena and relationships). To achieve this purpose, the integration is splitted in two major parts, as Figure 1 shows.

The syntactic translator is responsible to transform the input conceptual schemas into a canonical syntactic file (CSF) format. Then the semantic translator performs an algorithm based on an ontology (Guarino, 1998) and on similarity matching (Cohen, Ravikumar & Fienberg, 2003) techniques to unify the content heterogeneities, generating a canonical syntactic and semantic file (CSSF).

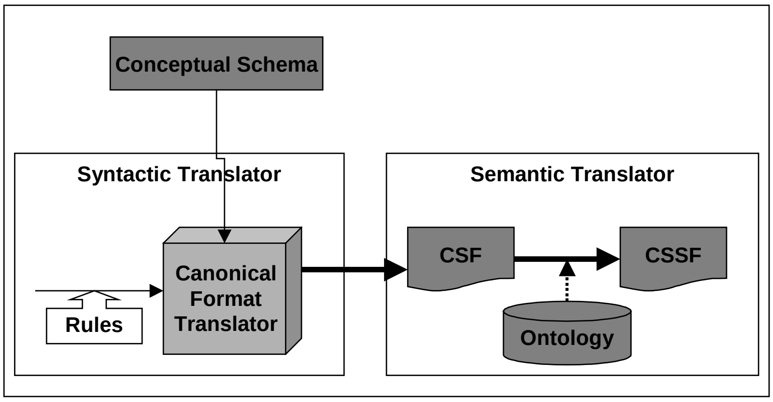

Figure 1: The integration architecture

Related Work

Fonseca et. al proposed an ontology-driven GIS architecture to enable geographic information integration. In that proposal, the ontology acts as a model independent system integrator (Fonseca et. al, 2002). Since it uses the hierarchical representation of ontologies to compare concepts, in many situations only a partial integration is possible, when finding the same super-concept of two specialized concepts.

Hakimpour and Timpf proposed the use of ontology in the resolution of semantic heterogeneities especially those found in Geographic Information Systems. In that work (Hakimpour & Timpf, 2001) the authors specify the ontological issues, present a little set of concepts in Description Logic (DL), and the concept's features that must be considered when solving heterogeneities: names, relations (attributes), and taxonomic relations. However, neither a methodology nor an algorithm are presented.

Stoimenov and Djordjevic-Kajan propose the GeoNis framework (Stoimenov & Djordjevic-Kajan, 2003) to reach the semantic GIS data interoperability. The use of ontologies was proposed as a knowledge base to solve semantic conflicts as homonyms, synonyms and taxonomic heterogeneities. The need of some type of syntactic integration is also pointed as an important task in order to unify the possibly different data models.

Even though most of the works are focused in the data itself while our work is focused in conceptual schemas, the idea behind is practically the same. However, none of the above cited works specifies an algorithm or methodology to search the ontology given a concept. Furthermore, no matching methods were described.

The semantic matching using the set theory was proposed by Tversky (Tversky, 1977). Rodriguez and Egenhofer propose an approach using ontologies and based on the properties and operations of the set theory to determine semantic similarity among entity classes from different ontologies (Rodriguez & Egenhofer, 2003). In that work, a similarity function determines similar entity classes by using a matching process over synonym sets, semantic neighborhoods and distinguishing features that are classified into parts, functions and attrinutes (Rodriguez & Egenhofer, 2003). The main difference from our approach to this one is the fact taht we do not use the set theory and also we measure similarity between concept's name if they are not in a synonym set.

Recent works in the ontology integration field present some techniques and methodologies to compare concepts and find correspondences. The H-Match (Castano et. al. Ferrara, Montanelli & Racca, 2004) is an algorithm for dinamically performing ontology matching in a multi ontology context. H-Match provides several matching models and metrics to work at different levels of depth, supporting interoperability in a flexible way (Castano, Ferrara, Montanelli & Racca, 2004).

THE SYNTACTIC INTEGRATION

To make the integration possible three requisites must be satisfied (Batini et. al, 1986):

- The conceptual schema of each source must be available;
- A canonical data model must exist with enough expressive power to describe all the models to be integrated;
- There must be semantic information in the schema;

Since the target of the integration proposed in this paper is of conceptual schemas, the first requisite is automatically satisfied. The second one is the target of this section and the last one is addressed in the semantic integration section.

While the entity-relational and the object-oriented models are standards in the traditional conceptual schemas design, the same does not occur with the GDB conceptual modeling. As each GIS system implements its own data model, there is not a *de facto* standard. This fact forces one to design a conceptual schema to a specific GIS architecture, what makes the interchange and reuse of the modeling almost impossible between different GIS systems. To make the conceptual design independent of GIS platform, some conceptual data models and *frameworks* have been proposed, such as GeoFrame-T (Rocha et. al., 2001), MADS (Parent et. al., 1999) and GeoOOA (Koesters et. al., 1997). However, even though these models are similar in their essentials, they are not easily interchangeable and, furthermore, none of them is accepted as a standard by a large community.

The Open GIS Consortium (OGC) developed a data model and format. In its current release, the Geography Markup Language (GML) version 3 (OGC, 2003) has constructors for representing almost every type of geographic phenomena, such as temporality, spatiality in both field and object views and networks. The GML is an XML based language and thus it is proper for information interchange. Thus, we adopted the GML 3 as our canonical data model.

As most of the GIS independent data models proposed are object-oriented, for the syntactic translation we developed a set of rules to convert the object-oriented constructors to GML tags. Our starting point was the comparison of a number of data models to identify common constructors. This set was then mapped into GML 3 (Hess & Iochpe, 2003). As the result of this phase of the integration process the CSF is generated.

The translation process consists of six rules, each one responsible to map a different object oriented constructor into a GML 3 element or set of elements: packages, classes, attributes, associations, and taxonomies. Additional rules were created for the geometric and temporal features.

We implemented tools for converting GeoFrame-T and MADS. As these models have their own case tools (in fact, GeoFrame-T uses the Rational Rose environment) and, furthermore, different structures, for each one a different program was written, implementing the general object-oriented rules specifically for that model. The GeoFrame’s one was written in Visual Basic for Applications (VBA) and the MADS’ in JAVA. Even though the models are syntactically different, the GML results were semantically equivalent for the same schema described using these two models.

THE SEMANTIC INTEGRATION

Only having the schemas described in the same data model (format) does not guarantee that the schemas are correctly unified. There may be some conflicts at the semantic level, which includes subjects related to the comprehension and due to the use of data in different applications and by different users with distinct interpretations of the geographic phenomena. Thus, the semantic converter module is responsible for handling and solving all sorts of semantic conflicts and heterogeneities.

It is a consensus that to perform the semantic integration in a semi-automate way, allowing the reuse of the modeled elements, the use of a Knowledge Organization System (KOS), such as an ontology, is mandatory (Hodge, 2000). The role of the ontology is similar to the role of the global conceptual schema proposed in the works of (Batini et al., 1986) and (Hayne et al., 1990). It also may be seen as a mediator as proposed by (Fonseca et al., 2002). Each one of the conceptual schemas to be integrated is compared with the ontology, and for each conflict found the system calculates a similarity measurement between the ontology's concept and the input conceptual schema's element.

Similarity measurement

The similarity of two concepts (classes or elements) can not be binary (1 to equivalent and 0 to not equivalent). To reach an accurate level of similarity between the concepts some mathematical formulas must be used to compute a value within 0 and 1 in the similarity measurement. To measure the similarity between two concepts, one from the input conceptual schema and the other from the ontology, we combine syntactic matching between strings (e.g., pairs of concept names as well as pairs of attributes names) and semantic matching (Hess & Iochpe, 2004). The syntactic matching has nothing to deal with the resolution of syntactic heterogeneities in GDB conceptual schemas. We assume here that the schemas are already described in the same data model. Thus, the syntactic considered here is in respect of concept names and attributes.

For the syntactic matching evaluation, a distance function is applied over a pair of strings, to determine the dissimilarity between them. In this work we adopted the Levenshtein distance (Cohen, Ravikumar & Fienberg, 2003), which returns the number of character changes needed to transform one string into another. The smaller this kind of dissimilarity (measured by an integer value) is, i.e., the less character changes needed, the more similar are the strings (Cohen, Ravikumar & Fienberg, 2003).

$$SimName(Cc,Co) = 1 - (Lev(Cc_{Name},Co_{Name})) \qquad (1)$$

where Cc and Co are, respectively, the conceptual schema's and the ontology's concept, and CC_{name} and CO_{name} are, respectively, the term used to nominate the conceptual schema's and the ontology's concepts.

The techniques to calculate the distance between two strings may be applied to acronyms and typing error cases, but no semantic issues are considered by these functions. Therefore, a correct semantic unification of concepts must be accomplished by a complementary technique that is capable of both detecting synonyms and considering the context in which those concepts exist.

Two semantic techniques are considered to compare a pair of concepts. The first one is the nearest neighbor (Holt, 2000), which calculates the similarity in terms of concept's attributes and is given by :

$$SimAt(Cc,Co)=\sum_{i=1}^{n} f(Cci,Coi)xWati \qquad (2)$$

where *n* is the number of attributes considered, *i* is the index of the attribute being processed, $f(Cc_i,Co_i)$ is the distance function between the attributes of the compared concepts (Levenshtein) and Wat_i is the weight of the i^{th} attribute in the ontology which is given by an adapted TF-IDF (Cohen, Ravikumar & Fienberg, 2003) formula:

$$W_{at}=1-(C_a/C) \qquad (3)$$

where Ca is the number of concepts having the attribute, and C is the total number of concepts. The more concepts have the same attribute, the less significant this attribute is.

Three types of relationships are considered for the similarity measurement of a pair of concepts. The first one is the taxonomic (IS-A) associations, and the others two are the aggregation and composition ones. The similarity in terms of the place in the hierarchy where each concepts is located is obtained by the formula:

$$SimHier(Cc,Co)=\frac{(\sum(Hier(Cc,Pc).Wt(c,p))}{Nhier(Cc,Pc))} \quad (4)$$

where Hier(Cc,Pc) is each one of the taxonomic relationships existing in both the conceptual schema and in the ontology. Wt(c,p) is the weight of the hierarchical relationship arc and Nhier(Cc,Pc) is the number of IS-A associations in both the ontology and the conceptual schema. The weight Wt(c,p) of a taxonomic arc is given by the following formula (Jiang & Conrath, 1997):

$$Wt(c,p)=\frac{(E)}{E(p)}.(\frac{d(p)+1}{d(p)}).(IC(c)-IC(p)) \quad (5)$$

where d(p) is the depth of the parent node (p) of the node corresponding to the concept being compared. E is the density of the whole ontology's hierarchy (the number of nodes it has). E(p) is the density of the taxonomy considering the node p as the root concept (the number of direct and indirect children it has). Finally, IC (Information Content) represents the amount of information the node has (Resnik, 1998), and its value is given by:

$$IC(c)=-log((\sum(1/sup(c))).1/N) \quad (6)$$

where sup(c) is the number of super classes (direct or indirect) the class c has, and N is the total number of concepts of the ontology. The more specialized a concept is, the more information it intrinsically possesses.

Finally, the aggregation and composition links are considered to calculate the similarity of two concepts, by the simple formula:

$$SimRel(Cc,Co)=(\sum(Rel(Cc,Co))/Rel(Cc)) \quad (7)$$

where Rel(Cc,Co) is each composition/aggregation link existing both in the ontology and in the conceptual schema and Rel(Cc) is the ones present only in the conceptual schema.

The final value of similarity is given by balanced sum of the similarities:

$$Sim(Cc,Co)=WN.SimName(Cc,Co)+WA.SimAt(Cc,Co)+ WH.SimHier(Cc,Co)+WR.SimRel(Cc,Co) \quad (8)$$

where WN, WA, WH and WR are the weights for names, attributes, hierarchies and relationships similarities, respectively.

The Sim(Cc, Co) value is calculated for every concept in the conceptual schema against each concept present in the ontology. The higher the Sim(Cc,Co) value is, the more similar the concepts are.

During this process of similarity measurement, the original conceptual schema is kept unaltered and a new, equivalent one is generated relying on the canonical semantic model expressed by the ontology. The ontology can also be updated during the algorithm execution depending on the similarity measurements carried out with every new GDB schema. Attributes and relations can be added to existing concepts, and even new concepts may be inserted. This update may be automatic, if the input concept does not have a minimum similarity with any ontology's concept or manual, performed by the expert, if he considers that the input concept is different from all the ones present in the ontology. Furthermore, even if the expert selects an existing concept, it may be updated if there are properties in the input concept missing in its correspondent in the ontology.

The ontology algorithm

Figure 2 shows the algorithm we developed in order to evaluate the degree of similarity of pairs of concepts (Hess, Iochpe & Oliveira 2004). It relies upon an ontology and can also be used to extend this ontology with both new terms (i.e., synonymous) and new concepts. Due to space limitations, we describe the algorithm at a high level of abstraction showing only a sequence of steps that can be used in searching as well as updating the ontology.

To minimize the intervention of an human expert, two parameters must be entered in each execution of the algorithm: the analysis threshold and the acceptance threshold. Only the concepts having similarity probability higher than the specified analysis threshold are considered. If no candidate reaches the threshold, the input concept is considered as not existing in the ontology and should be added as a new concept. If one ore more of the ontology candidates have similarity probability higher than the specified by the acceptance threshold, the one with the higher value is considered equivalent (a synonym) to the input concept.

To ensure the correct operation of the algorithm, it is necessary that every input conceptual schema has an associated metadata, specifying in which language the modeling is based.

Step 0 – Schema translation into the ontology's language: If the ontology's language is not the same as the one indicated by the conceptual schema's metadata, the last has to be translated, aided by a dictionary.

Step 1 – Search concept's name in the ontology: If the concept's name or one of its synonyms or acronyms is found in the ontology, go to step 2. Else go to steps 4, 5 and 6, in parallel.

Step 2 – Search concept's structure in the ontology: Once the term which nominates the concept is found in the ontology, its structure is compared against the ontology, attribute by attribute. The algorithm verifies if there is a one to one correspondence between each input concept's attribute and the ontology concept's attribute. If the equivalence is complete go to step 3. If there are differences in at least one of the attributes, go to step 5.

Step 3 – Tests if it is the last concept: If the current concept is the last one of the schema, go to the end. Else go back to step 1 to processes of the next concept.

Step 4 – Calculate the similarity of the term that nominates the concept: The similarity of the input concept's name is compared against the name of each ontology's concept. Go to step 7

Step 5 – Calculate concept's structural similarity: The input concept's structural similarity is calculated, in terms of its attributes against each one of the ontology's concept. Go to step 7.

Step 6 – Calculate relation similarity: The input concept's relation similarity is calculated, in terms of aggregation and composition associations, and also in terms of taxonomic (IS-A) relations against each one of the ontology's concept. Go to step 7.

Step 7 – Sum of the similarities: Based on some method of balance, the structural similarity, the name similarity and the relation similarity of the input concept are summed, resulting in the similarity probability. This calculation is performed for each ontology's concept. Go to step 8.

Step 8 – Verify threshold: Check if the similarity value of the concept with the highest similarity probability. If it is lower than the analysis threshold go to step 12. If the similarity value is greater than the acceptance threshold and there are no other candidates with similarity probability higher than the highest similarity value go to step 11. Otherwise, go to step 9.

Step 9 – Show candidates: Present each found candidate, with its balanced similarity probability. They are displayed ordered, with the ones with higher similarity first. Go to step 10.

Step 10 – Term selection: At this point the domain expert intervention is necessary. He selects the concept he judges as the most equivalent to the input schema's concept. If an ontology's existing concept is selected, go to step 11. If the expert decides that the input concept does not have an equivalent in the ontology go to step 12.

Step 11 – Update an existing concept: Depending on which step called this step, a distinct action is performed to update the ontology. This action can be the addition of a new synonym or acronym to an existing term, the addition of a new attribute to an existing concept's structure, or the creation of a new relation between two existing concepts. Go back to step 3.

Step 12 – Addition of a new concept to the ontology: A new concept is added to the ontology, with all its attributes and relations. Go back to step 3.

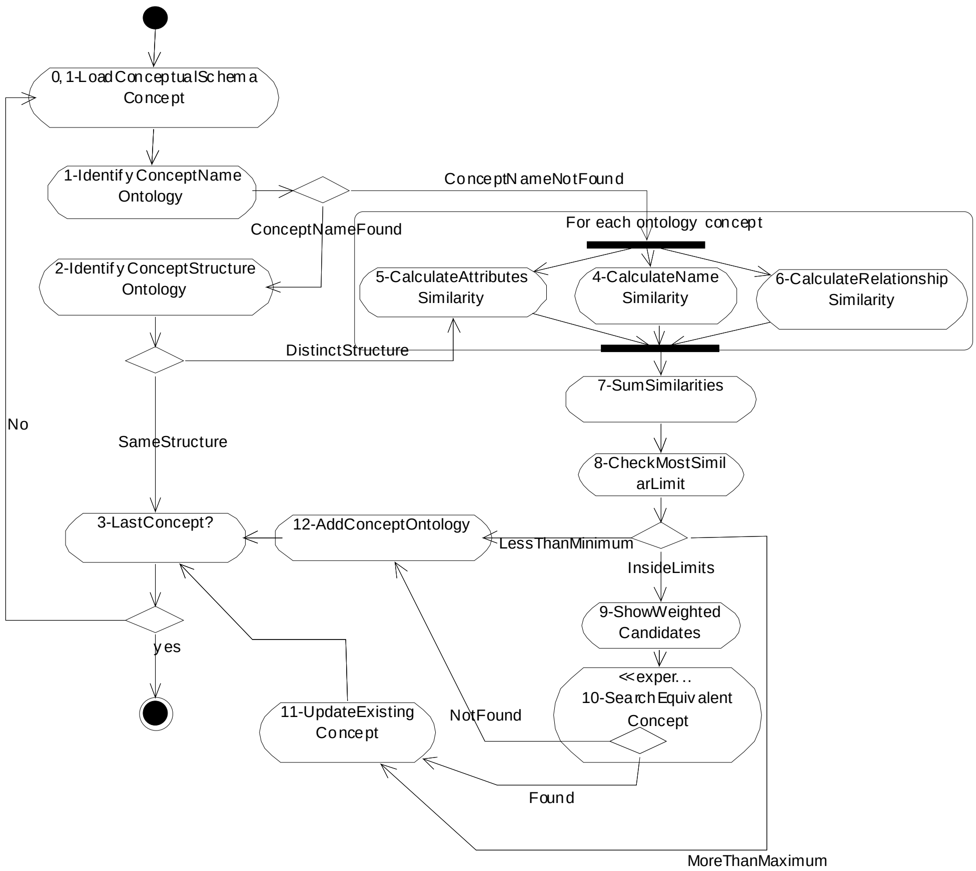

Figure 2: The ontology search and update algorithm

The human intervention in the resolution of the conflict is practically mandatory in the identification of correspondence process between different schemas. At most, what can be reached is that the ontology suggests the best solutions based on similarity and probability calculus.

CASE STUDY

The syntactic step intends to convert the schemas into a canonical format, the Geographic Markup Language (GML). For each one of the data models used (MADS, GeoFrame and OMT-G) a set of translation rules was built and a prototype was developed to convert MADS and GeoFrame to the GML. The results were very correct, since the same schema modeled based on the different data models generated the same GML elements. The files were not exactly the same just because the order of the elements and type declarations were different.

Having all the conceptual schemas described in GML, the semantic phase takes place. In order to test the algorithm as well as the similarity matching formula proposed, a geographic ontology called ontoGeo was built. The best choice to cope with the goals above was the one of a domain ontology. A side effect of this decision is that any other geographic application may use the concepts stored in this ontology. The concepts contained in it are in the domain of the physical, natural phenomena (basic cartography) such as hydrology, relief, and growth. Some concepts of the infra-structure domain are also present, especially those of the transportation theme.

Based on the geographic data and the types of relationships between them, the concepts may be divided into five different categories: classes, spatial representation (geometry), temporal representation, network associations, and simple relationships. To describe the first four categories, we used the classes of the ontology. Associations of concepts as well as concept attributes are represented by slots (attributes) in the ontology. Slots were used also to associate concepts to one another regardless whether they belong to the same or to different categories. The concepts described in ontoGeo were taken from several conceptual schemas that relied on different data models and were designed by different people. It is worth nothing that ontoGeo is not intended to be a complete geographic ontology. It serves as a starting point to integrate GDB conceptual schemas. In the case a concept belonging to an input schema is not found in the ontology, ontoGeo may be updated with the insertion of these new concept as predicted in the presented algorithm.

For the execution of the algorithm that searches and may update the ontology, we used only a subset of ontoGeo's concepts. The chosen sub-domain was hydrology, because of the availability of conceptual schemas for the test that presented this theme in their design.

We executed the experiments considering three different types of conceptual schemas. The first type represents those schemas modeling only classes and taxonomic relations. Schemas modeling classes with attributes and taxonomic relations were considered as belonging to a second type. The last type represents schemas that can have both classes with and without attributes, besides hierarchies. Furthermore, when a GDB schema of the third type was processed, ontoGeo had already been updated on the basis of schemas of the other two types.

For each schema the similarity matching algorithm was executed twice, each time with a different set of values for the weights WN, WA, WH, and WR (Table 1). In the following, each one of these sets is called a different scenario.

Table 1: The case study scenarios

Input Conceptual Schema	**WN**	**WA**	**WH**	**WR**
Classes and Hierarchies	0.25	0.25	0.25	0.25
Classes and Hierarchies	0.70	0.00	0.30	0.00
Classes, Attributes and Hierarchies	0.25	0.25	0.25	0.25
Classes, Attributes and Hierarchies	0.45	0.35	0.20	0.00
Classes, Attributes, Hierarchies, Ontology Updated.	0.25	0.25	0.25	0.25
Classes, Attributes, Hierarchies, Ontology Updated.	0.50	0.25	0.25	0.00

For all three GDB schemas processed, in the first scenario all weights were given the value of 0.25. In the second scenario for each schema the WN, WA, WH, and WR assumed different values depending on the characteristics of the input conceptual schema. Furthermore, the acceptance threshold was fixed in 0.75 (75%), while the analysis threshold was set to 0.4 (40%).

In the first round of execution, no equivalent concept in ontoGeo was found for any of the processed concepts of the input schema. In this case, SimAt and SimRel were always zero, and thus the similarity had at most the value of 0.5. As we increased the values of both WN and WH, and decreased the values of WR and WA down to zero, the algorithm showed very good results, with 100% of accuracy in automatically matching equivalent concepts with more than 75% of similarity, and 60% of precision in finding equivalent concepts in ontoGeo within 40% and 75% of similarity.

The second conceptual schema that was evaluated against ontoGeo contains taxonomies, classes and attributes. In a first execution of the algorithm, the results were 100% correct in the cases where similarity fell within the threshold limits. Again, no match was indicated as having more than 75% of similarity because no aggregation or composition associations were modeled. In a second execution, by increasing the values of WN and WA as well as decreasing the value of WH a little and setting WR to zero, the results were the same and also a correct automatic matching was obtained.

The third conceptual schema contains taxonomies, classes and attributes. It was processed by the similarity matching algorithm after the ontology had been updated. As most of the constructs of this third schema were already stored as concepts in the ontology, the results of the two executions (i.e., one for each different scenario) were almost the same.

Summaryzing the case study executions, the success on automatic matches oscillated between 60% and 80% and the success in presenting candidates to the expert oscillated between 85% and 100%.

CONCLUSIONS

This work aimed to address an important issue for the interoperability of geographic data and conceptual schemas, which is the integration of different schemas. Input schemas must be syntactically as well as semantically unified as to produce good results. Especially for GDB conceptual schemas, the integration plays a very important role since it handles heterogeneities in terms of data models (syntax) and in terms of concepts (semantics). It was made clear that the use of an ontology associated to a canonical data model helps to enhance the schema unification process.

The more detailed the conceptual schema is, the more precise is the algorithm output. When only classes were modeled a considerable number of matching candidates were found in the ontology for each one of the concepts of the input GDB schema, as only the concept's name differed. When the elements were modeled in a complete taxonomy and with a number of attributes the ontology's concepts had to match more requisites, producing less, but more accurate, candidate matches.

In the case study we report here, the main goal was to determine a good scenario where the weights for the different types of similarity formulas express their relative importance within the global similarity formula (8). A next step should be to investigate the best set of values for the set of weights concerning both the similarity matching algorithm and the ontology ontoGeo. It is possible, though, that one comes up with a set of good scenarios, each one of which being the best fit for a certain type of GDB schema. Thats is because there is not a perfect combination of values for the set of weights composed by WN, WA, WR and WH. The relations among these depend on the characteristics of the input conceptual schema being processed. In this work we did not test all different types of schema that may exist.

An important point to be considered for a future work is the fact the algorithm is not yet capable to solve heterogeneities in terms of different data model constructors used to model the same concept of the real world. For example, if a concept in represented as a class in one schema and as an attribute in another, the current algorithm does not match one to another.

BIBLIOGRAPHY

Batini, C. et. al. A Comparative Analysis Of Methodologies For Database Schema Integration. In ACM Computing Surveys, v.18, n.4, p.323-364, 1986.

Burrough, P. A.; McDonnel, R. A. Principles of Geographical Information Systems. Great Britain: Oxford university Press, 1997.

Castano, S.; Ferrara, A.; Montanelli, S.; Racca, G. Semantic Information Interoperability in Open Networked Systems. Proc. of Int. Conference on Semantics of a Networked World (ICSNW), in cooperation with ACM SIGMOD. pp. 215-230. Paris, France, 2004.

Cohen, W. W.; Ravikumar, P.; Fienberg, S. E. A comparison of String Metrics for Matching Names and Records. Proc. of IJCAI 2003 – Workshop on Information Integration on the Web. Acapulco, Mexico, august, 2003.

Fonseca, F. T., Egenhofer, M. J., Agouris, P. and Câmara, G. Using Ontologies for Integrated Geographic Information Systems. In: Transactions in GIS 6(3), 2002.

Fonseca, F. et. al. Bridging Ontologies and Conceptual Schemas in Geographic Information Integration. GeoInformatica. v.7, n.4, p.355-378. Kluwer Academic Publishers, 2003.

Guarino N., Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. In M.T. Pazienza (ed.). Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology. Springer Verlag: 139-170, 1998.

Hakimpour, F.; Timpf, S. Using Ontologies for Resolution of Semantic Heterogeneity in GIS. In 4th AGILE Conference on Geographic Information Science. Brno, Czech Republic. April, 2001.

Hayne S. and Ram, S. Multi-User View Integration System (MUVIS): An Expert System for View Integration. Proc. of the 6th International Conference on Data Engineering, p. 402-409, 1990.

Hess, G. and Iochpe, C. Using the GML for the Identification of GDB Analysis Patterns Candidates. Proc. of the V Brazilian Symposium on GeoInformatics. Campos do Jordão, Brazil, 2003.

Hess, G. and Iochpe, C. Applying Ontologies in the KDD Pre-Processing Phase. Proc. of 16th International Conference on Software Engineering and Knowledge Engineering (SEKE'04). Banff, Canada. 2004.

Hess, G. N.; Iochpe, C.; Oliveira, J. P. M. de. Analysis Patterns for the Geographic Database Conceptual Schema: An Ontology Aided Approach. Proc. of the International Conference on Semantics of a Networked World (ICSNW). pp. 323-324. Paris, France, 2004.

Hodge, G. Knowledge Organization Systems: An Overview. In System of knowledge Organization for Digital Libraries: Beyond Traditional authority files. 2000.

Holt, A. Understanding environment and geographical complexities trough similarity matching. In Complexity International, n.7, 2000.

Jiang, J. and Conrath, D. Semantic Similarity Based in Corpus Statistics and Lexical Taxonomy. Proc of International Conference Reasearch in Computational Linguistics. Taiwan, 1997.

Köesters, B. et. al. GIS-application development with GeooOA. International Jouranl of Geographic Information Science. London, England, 1997.

OpenGIS Consortium. Geography Markup Language (GML) 3.0. Open GIS Implementation Specification, 2003. Available in http://www.opengis.net. Last access in december 2003.

Parent, C. et al. Spatio-temporal conceptual models: data structures+space+time. Proc.7th ACM GIS, Kansas City, USA, 1999.

Resnik, P. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. Journal of Artificial Intelligence Research, n.11, p. 95-130, 1998.

Rocha, L. V. et. al. "GeoFrame-T: A Temporal Conceptual Framework for Data Modeling". In: ACM Symposium on Advances in GIS. Atlanta, USA, 2001.

Rodriguez, M. A.; Egenhofer, M. J. Determining Semantic Similarity Among Entity Classes from Different Ontologies. IEEE Transactions on Knowledge and Data Engineering, 15(2): 442-456. 2003.

Stoimenov, L.; Djordjevic-Kajan, S. Realization of GIS Semantic Interoperability in Local Community Environment. Proc. of the 6th AGILE Conference on Geographic Information Science. Lyon, France, 2003.

Tversky, A. Features of Similarity. Psychological Review, 84 (4): 327-352. 1977.