Matching Spatial Ontologies
A Challenge of Formalization

Sana Châabane and Faïez Gargouri
MIRACL laboratory,University of Sfax- Tunisia, Route de Tunis km 10, Cité El Ons Sfax 3021, Tunisia

Keywords: Ontology Matching, Spatial Ontology, Merging, Mapping, Semantic Matching, Topologic Matching, Geometric Matching.

Abstract: Ontology matching can be defined as the process of defining a set of functions for specifying correspondences between candidate concepts in order to discover similarities between two ontologies; it can be processed by exploiting a number of different techniques. In this paper, *w*e present an approach of merging spatial ontologies which consists of three processes: “matching process”, “mapping process” and “merging process” and we focus on the matching process. Indeed we consider three kind of matching: semantic matching, topologic matching and geometric matching. For each type of matching, we formally define functions for specifying correspondences between candidate concepts.

1 INTRODUCTION

The problem of heterogeneity has been addressed in different research communities, particularly in the context of databases cooperation (Bin et al., 2003) and (Giunchiglia and Shvaiko, 2003) but also in order to share correctly and efficiently all the knowledge contained in different ontologies (Noy and Musen, 2000). Different processes are used to share the knowledge contained in several ontologies: integration, merging, alignment. These processes differ depending on the desired results, available information, the level of integration, etc.

Ontology matching takes an important role in the process of ontology integration and merging with the purpose of establishing semantic relationships between two ontologies. In general, ontology matching can be defined as the process of discovering similarities between two ontologies (Predoiu et al., 2006). It determines the relations holding two sets of entities that belong to two discrete ontologies (Ehrig and Sure, 2004). In other words, it is the process of finding a corresponding entity in the second ontology for each entity (for example, concept, relation, attribute) in the first ontology that has the same or the closest intended meaning. This can be achieved by analyzing the similarity of the entities in the compared ontologies in accordance with a particular metric (Ehrig and Sure, 2004) and (Interop, 2004).

We are interested in our work, in merging spatial ontologies. Merging creates a single coherent ontology; different ontologies about the same domain are merged into one that "unifies" all (Noy and Klein, 2003). Indeed, spatial information, diverse in nature, is a specific case of heterogeneity, because of the multitude of data handled sources. A spatial object is an object modeling a real world phenomenon, particularly in describing one or more locations on the globe surface. A spatial object is described by semantic data (its name, its nature, its appearance, its various characteristics ...) and by geometric data (its position on the surface). We consider that a spatial ontology consist of spatial concepts, semantic relations and spatial relations (Sana et al., 2013). Semantic relations are those supported by UML (Xu et al., 2008; Andy et al., 1998; Ruth et al., 1997; OMG, 1997) and (Rumbaugh et al., 1998), that are: generalization, aggregation, composite and simple association with a name. Spatial relations are of three types: metric relations classified into two types: Distance relations that express a distance with a value and a unit of measure, and approximate relations that express an approximate distance between two spatial objects. Considered approximate relations are: {in-side, near, beside, nigh}. The second type of considered spatial relations is directional relations that model the nine cardinal positions of a spatial object and express the position of a spatial object versus another.

Châabane S. and Gargouri F..
Matching Spatial Ontologies - A Challenge of Formalization.
DOI: 10.5220/0004623603550360
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2013), pages 355-360
ISBN: 978-989-8565-81-5

Directional relations are defined throughout the DIRECTION set: {North, South, East, West, North East, North West, South East, South West} formally presented in section four. Finally, topological relations are those defined in (Clementini et al., 1993).

Spatial ontologies have shown that respecting the independence between the conceptual level and external level, it is possible to provide different views of ontology. The problem of heterogeneity of spatial ontologies is more complex than that of other domain ontologies, because it is necessary to take into account the spatial aspects of concepts and relations. This field of study is not yet well explored by researchers, thing that incites us to propose an approach of merging spatial ontologies. In this paper, we focus on the matching process of this approach.

This paper is organized as follows: second section presents an overview on techniques of matching ontologies. Third section details our approach of merging spatial ontologies. In the fourth section, we detail matching process by presenting informal and formal definitions of matching functions. We conclude this paper by conclusion and future work we intend to achieve.

2 TECHNIQUES OF MATCHING ONTOLOGIES

Many works has been developed in the field of ontologies matching based on basic techniques of specification of methods for calculating semantic distances between concepts and tools that are more complete and integrating these methods into their processes of matching. To provide a common conceptual basis, researchers have started to identify different types of ontology matching techniques and propose classifications to distinguish them. In (Abels et al., 2005) a classification is defined that consists of nine matching techniques based on existing literature studies. Another classification of these techniques of matching is proposed in (Euzenat and Shvaiko, 2007). This classification is based on the classification proposed by (Rahm and Bershtein, 2001) and considers other criteria for comparing matching approaches. We detail in what follows, matching techniques presented by (Euzenat and Shvaiko, 2007).

2.1 Linguistic Techniques

These are all techniques used to evaluate the similarity between two concepts based on their names and names of their properties. The common characteristic of linguistic techniques is to produce a measure of similarity between two strings. These results can be obtained by syntactic, lexical and semantic techniques. The three types of linguistic techniques may be used in combination.

2.2 Contextual Techniques

These techniques are based on the fact that the meaning of a concept is strongly linked to context. Indeed, they are taken to assess the similarity between concepts by analyzing their contexts. The context is represented by different structures describing concepts such as properties of the concept or semantic relations between concepts of the ontology. There are many different techniques for the assessment of contextual similarity, there are those that are based on heuristics metrics and those based on probabilistic reasoning.

2.3 Combined Techniques

Tools for identifying matches are not based on a single technique but rather on a combination of different techniques in order to obtain a comprehensive measure of similarity. Once this measure is defined, it must devise a mechanism to eliminate results deemed irrelevant.

2.4 Extensional Techniques

These techniques are based on the analysis of ontology instances with statistical methods, and probabilistic learning.

2.5 Techniques based on Neighborhood Structures of Concepts

Concepts comparison may be realized on the concept's name itself and its neighbors in the ontology with inheritance hierarchies or relationships of concepts (domains of departure or arrival, multiplicity, etc…).

These techniques are relevant to domain ontologies; namely spatial ontologies which have specific characteristics related to the spatial aspect of concepts and spatial relations. We must therefore, take into account these characteristics in the merging process of spatial ontologies.

In the next section, we present our approach of merging spatial ontologies.

3 APPROACH OF MERGING SPATIAL ONTOLOGIES

The proposed approach takes as input two spatial ontologies called candidate ontologies, and provide as a result a single ontology called global ontology. It consists of three phases. The first phase is to apply a matching process between candidate ontologies. "Matching" is the process of defining a set of functions for specifying correspondences between candidate concepts (Shvaiko and Euzenat, 2008). A matching function is a binary relation between two spatial concepts. The second phase is "mapping" which consist to find correspondences between candidate concepts referring to the matching functions definitions. The third phase is "merging" which consist to build the resulting ontology based on merging rules which is spatially and semantically richer. "Figure 1" presents our approach of merging spatial ontologies.

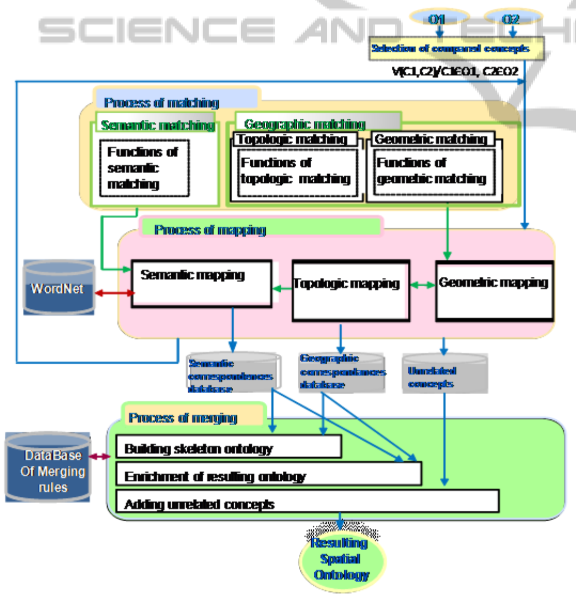

Figure 1: Approach of merging spatial ontologies.

In this paper, we focus on the matching process. The following section presents in detail the various steps of the matching process.

4 THE PROCESS OF MATCHING

The first phase of our approach is the matching process. This process is to define relations of correspondence between spatial concepts representing the same real-world phenomena, but from two different data sets. This process involves a large number of techniques and parameters that depend heavily on the modeling of geographic information and also involves semantic aspects. We propose three types of matching:

The "semantic matching" is to match concepts using their semantic characteristics. The "topological matching" uses topological relations between concepts for matching. Finally, the "geometric matching" consists to match concepts using their geographical characteristics, including their location and their graphic shapes.

These three types of matching can be used separately or in addition to one another. In what follows, we define the matching functions adopted by our approach. For each matching function, an informal definition and a formal definition are given. To formally define the matching functions, we must first introduce the components of a spatial ontology (Sana et al., 2013). Indeed, we consider a spatial ontology consists of spatial concepts and semantic and spatial relations. A spatial concept is characterized by its name of string, its location and its graphic shape. We define the graphical shapes of a spatial concept as Point, Line or Polygon. A *Point* is characterized by an x and y coordinates of integer. A *Line* is characterized by the properties ds: start of section and fs: end of section of Point, a height of integer and a direction. A *Polygon* is characterized at least three extremities of Point: e1, e2 and e3.

To formally write matching functions, we take the following parameters:

C1, C2, C3: spatial concepts.

P1, P2: two spatial concepts of Point graphic shape;

G1, G2: two spatial concepts of Polygon graphic shape;

L1, L2: two spatial concepts of Line graphic shape;

T1, T2, T3: terms of strings.

A(C1): The set of C1 attributes (set of strings).

A(C2): The set of C2 attributes (set of strings).

A(C3): The set of C3 attributes (set of strings).

4.1 The Semantic Matching

The semantic matching aims to express more semantics of concepts and to show semantic links between concepts of candidate ontologies; thereafter semantically enrich the resulting ontology. In our approach, semantic matching is based on the calculation of similarities between candidate concepts. In fact, we adopted a combinatorial technique for similarity evaluation between two concepts based on their names and properties. A combinatorial technique means that it is a combination of a syntactic technique which is "edit distance "ed" (Levenshtein, 1996) which represents the minimum insertion, deletion or substitution required to transform one string T1 into another T2; a lexical technique defined by (Maedche and Staab, 2002) and a semantic technique using an external resource (Miller, 1995) which is the knowledge base "Wordnet". After a detailed study on the nature of concepts of spatial ontologies, we define the following functions of semantic matching using description logic.

Rule 1 : The semantic-Identity relation is written ***Idsem***(C1, C2) is satisfied between two spatial concepts C1 and C2 if and only if C1 name is syntactically equal to C2 name and the set of attributes of C1is equal to the set of attributes of C2. ***Idsem***( ) function is defined formally as follows:

$$Idsem := \left\{ (C1, C2) \middle| \begin{array}{c} C1 \in O1 \wedge C2 \in O2 \\ \wedge\, synt(C1.Name, C2.Name) = \\ Eens\big(A(C1), A(C2)\big) \end{array} \right. \quad (1)$$

The function ***Idsem***( ) is symmetric and transitive.

To define the ***Idsem***( ) function, we define the functions ***synt***( ) and ***Eens***( ) whose definitions are given below:

Rule 2: The syntactic-equality relation written ***synt***(T1, T2) is satisfied between two terms T1 and T2 if and only if the edit distance "ed" between the two terms is equal to zero. Then, we write:

$$synt(T1, T2) = \begin{Bmatrix} 1\ Si\ ed(T1, T2) = 0 \\ 0\ Si\ ed(T1, T2) \neq 0 \end{Bmatrix} \quad (2)$$

The function ***synt***( ) is symmetric and transitive.

Rule 3: The Equality-sets relation written ***Eens***(A(C1),A(C2)) is satisfied between two sets of terms if and only if for any term T1 belonging to the first set there exists a term T2 belonging to the second set where ***synt***(T1,T2)=1, and vice versa. Then, we write:

$$Eens := \left\{ \begin{array}{c} A(C1), A(C2)/ \\ \left\{ \begin{array}{c} \forall a1 \in A(C1)\ \exists \\ a2 \in A(C2)\ /synt(a1, a2) = 1 \end{array} \right\} \\ \wedge \\ \left\{ \begin{array}{c} \forall a2 \in A(C2)\ \exists \\ a1 \in A(C1)\ /synt(a2, a1) = 1 \end{array} \right\} \end{array} \right\} \quad (3)$$

The function ***Eens***(A1, A2) is symmetric and transitive.

Rule 4: The Equivalence relation written ***Eq***(C1, C2) is satisfied between two concepts C1 and C2 if and only if the name of C1 is semantically or linguistically equal to C2 name and all the attributes of C1 admit linguistic inclusion or synonymic inclusion in the set of C2 attributes. Then, we write:

$$Eq := \left\{ \begin{array}{c} (C1, C2) | C1 \in O1 \wedge C2 \in O2 \wedge \\ \begin{pmatrix} synonyme(C1.Name, C2.Name) = 1 \\ \vee\ ling(C1.Name, C2.Name) = 1 \end{pmatrix} \wedge \\ (\begin{pmatrix} Incling\big(A(C1), A(C2)\big) \vee \\ Incling\big(A(C2), A(C1)\big) \end{pmatrix} \vee \\ \begin{pmatrix} Incsyn\big(A(C1), A(C2)\big) \vee \\ Incsyn\big(A(C2), A(C1)\big) \end{pmatrix}) \end{array} \right\} \quad (4)$$

The function ***Eq***( ) is symmetric and transitive.

To define ***Eq***( ) function, we define the functions: ***Incsyn***(), ***Incling***( ), ***synonyme***( ) and ***ling***( ) whose formal definitions are the following:

The Synonymy relation written ***synonyme***(t1,t2) is satisfied between two strings if and only if the terminological resource "Wordnet" generates a synonymy relation between t1 and t2.

The function ***synonyme***( ) is symmetric and transitive.

Rule 5: The Synonymic-Inclusion relation written ***Incsyn***(A(C1),A(C2)) is satisfied between two sets of strings if and only if every element of E1 belongs to E2 itself or its synonym.

$$Incsyn := \left\{ \begin{array}{c} A(C1), A(C2) | \forall a1 \in A(C1), \exists a2 \in A(C2) \\ /synt\ (a1, a2) = 1 \vee \\ synonyme(a1, a2) = 1 \end{array} \right\} \quad (5)$$

The function ***Incsyn***(A(C1), A(C2)) isn't symmetric but is transitive.

Rule 6: The linguistic-equality relation written ***ling***(t1, t2) is satisfied between two terms if and only if the lexical technique SM (T1, T2) ≥ 60%. "SM" technique (String Matching) (Maedche and Staab, 2002) uses the Levenshtein edit distance for the similarity calculation; it takes as input two lexical entries and returns a similarity degree between 0 and 1, where 1 means perfect similarity and 0 otherwise. We set a threshold of similarity: s = 60%.

$$Ling := \left\{ T1, T2 \middle| \begin{array}{l} 1\ si\ SM(T1, T2) \geq 60\% \\ 0\ \text{si}\ SM(T1, T2) \leq 60\% \end{array} \right\} \quad (6)$$

The function ***ling***( ) is symmetric and not transitive.

Rule 7 : Linguistic-Inclusion relation written ***Incling***(A(C1),A(C2)) is satisfied between two sets of strings if and only if every element of A(C1) belongs, itself or its linguistic equivalent, to A(C2). Then, we write:

$$Incling := \left\{ \begin{array}{c} A(C1), A(C2) | \forall a1 \in A(C1), \exists a2 \in A(C2) \\ /\ synt\ (a1, a2) = 1 \\ \vee\ ling(a1, a2) = 1 \end{array} \right\} \quad (7)$$

The function ***Incling***( ) isn't symmetric but is transitive:

4.2 The Topologic Matching

This type of matching is using topological relations to match concepts; such matching is using to spatially enrich the resulting ontology.

Rule 8:

$$Inclusion(P, L1) \wedge Inclusion(L1, L2) \Longrightarrow Inclusion(P, L2) \tag{8}$$

Rule 9:

$$Extremity(P, L) \wedge Rencontre(L, G) \wedge L \cap G = P \Rightarrow Connexion(P, G) \tag{9}$$

Rule 10:

$$Inclusion(P, G1) \wedge Inclusion(G1, G2) \Rightarrow Inclusion(P, G2) \tag{10}$$

Rule 11:

$$Adjacence(G1, G2) \wedge Inclusion(P, L) \wedge G1 \cap G2 = L \Rightarrow Connexion(P, G1) \wedge Connexion(P, G2) \tag{11}$$

Rule12:

$$Connexion(L, G) \wedge Inclusion(P, L) \wedge L \cap G = P \Rightarrow Connexion(P, G) \tag{12}$$

Rule 13:

$$Inclusion(L, G) \wedge Inclusion(P, L) \Longrightarrow Inclusion(P, G) \tag{13}$$

Rule 14:

$$Connexion(P, G1) \wedge Inclusion(G1, G2) \Rightarrow Inclusion(P, G2) \tag{14}$$

Rule 15:

$$Rencontre(L1, L2) \wedge Inclusion(L1, G) \wedge L1 \cap L2 = P \Longrightarrow Inclusion(P, G) \tag{15}$$

Rule 16:

$$Inclusion(P1, L1) \wedge P \in \{L1.ds, L2.fs\} \Rightarrow Extremity(P1, L1) \tag{16}$$

Rule 17:

$$Extremity(P1,L1) \Rightarrow Inclusion(P1,L1) \tag{17}$$

Rule 18:

$$Jonction(L1, L2) \wedge Extremity(P1, L2) \wedge L1 \cap L2 = P1 \Rightarrow Meet(L2, L1) \tag{18}$$

Rule 19:

$$Jonction(L1, L2) \wedge Extremity(P1, L1) \wedge L1 \cap L2 = P1 \Rightarrow Meet(L1, L2) \tag{19}$$

4.3 The Geometric Matching

The geometric matching uses graphic shapes of concepts and their locations for matching.

Rule 20: The Spatial-Identity relation written ***Idspa***(C1,C2) is satisfied between two spatial concepts C1 and C2, if and only if they have the same graphic shapes and the same locations. Then, we write:

$$IDspa := \left\{ (C1, C2) \middle| \begin{array}{c} C1 \in O1, C2 \in O2 \wedge \\ \begin{pmatrix} \mathrm{C1.FG = point} \wedge \\ \mathrm{C2.FG = point} \wedge \\ ((\mathrm{C1.X = C2.X}) \wedge \\ (\mathrm{C1.Y = C2.Y}) \end{pmatrix} \wedge \\ \begin{pmatrix} C1.FG = Line \wedge \\ C2.FG = Line \wedge \\ C1.ds = C2.ds \wedge \\ C1.fs = C2.fs \end{pmatrix} \wedge \\ \begin{pmatrix} C1.FG = Polygon \wedge \\ C2.FG = Polygon \wedge \\ \forall E1, E2, E3 \subset G1.E, \\ \exists \{E'1, E'2, E'3\} \subset \\ G2.E \end{pmatrix} \end{array} \right\} \tag{20}$$

The function ***Idspa( )*** is symmetric and transitive.

Rule 21: The Identity relation written ***Identity***(C1,C2), is satisfied between two spatial concepts C1 and C2 if and only if they admit ***Idsem***(C1,C2) and ***Idspa***(C1,C2) relations between them. Then, we write:

$$Identity := \left\{ \begin{array}{c} \mathrm{C1, C2} | C1 \in O1, C2 \in O2 \wedge \\ Idsem(C1, C2) \wedge Idspa(C1, C2) \end{array} \right\} \tag{21}$$

The function ***Identity( )*** is symmetric and transitive.

5 CONCLUSIONS

In this paper we defined the process of matching ontologies and we presented a survey of matching techniques. Then, we presented our approach of merging spatial ontologies which consists of three processes: "matching process", "mapping process" and "merging process". We focused on the "matching process" and presented formal definitions of matching functions. These functions serve as input to the mapping process in order to infer relations between concepts of candidate ontologies.

In future work, we intend to define algorithms of mapping process and rules of merging process. In a second step, we intend to build an automatic tool for merging spatial ontologies.

REFERENCES

A. Maedche, Steffen Staab., 2002. "Measuring Similarity between Ontologies", *Proceedings of the European Conference on Knowledge Acquisition and Management (EKAW)*.

Abels, S., Haak, L., Hahn, A., 2005. "Identification of Common Methods Used for Ontology Integration Tasks". In: *1stInternational ACM Workshop on Interoperability of Heterogeneous Information Systems (IHIS)*, pp. 75-78. Bremen, Germany.

Andy Evans, Robert France, Kevin Lano, Bernhard Rumpe, 1998. "Developing the UML as a formal modelling notation". In Pierre-Alain Muller and Jean Bezivin, editors, *Proceedings of UML'98 International Workshop*, Mulhouse, France, pages 297–307.

Bin HE, Kevin Chen-Chuan CHANG, Jiawei HAN, 2003. *"Automatic Complex Schema Matching across Web Query Interfaces: A Correlation Mining Approach"*. Technical Report UIUCDCS-R-2003-2388, Department of Computer Science.

Eliseo Clementini, Paolino Di Felice, Peter van Oosterom, 1993. "A Small Set of Formal Topological Relationships Suitable for End_User Interaction". *In proceedings of the Third International Symposium on Advances in Spatial Databases*: SSD'93, Pages 277-295.

Ehrig, M., Sure, Y., 2004. "Ontology Mapping – An Integrated Approach". In: Bussler, C., Davies, J., Fensel, D., Studer, R. (eds.) ESWS 2004. LNCS, vol. 3053, pp. 76-91. *Springer*, Berlin/Heidelberg

Euzenat, J. and Shvaiko P., 2007. *"Ontology matching techniques"*. In ontology Matching, springer verlag, pp 61-149.

G. A. Miller., 1995. *"WordNet: A lexical database for english"*. Communications of the ACM (CACM), 38(11):39–41.

Giunchiglia F. and P. Shvaiko., 2003. *"Semantic matching"*. Knowledge engineering review, 18(3).

Interop., 2004. *"Ontology Interoperability"*. State of the Art Report (SOA), WP8ST3 Deliverable, IST-508011.

Levenshtein, I. V., 1996. *"Binary codes capables of corrections, deletions, insertions and reversals"*. Soviet Physics-Doklady, 10(8):707_710.

Noy N.F. and Klein M., 2003. *"Ontology evolution: not the same as schema evolution"*. Knowledge and Information Systems.

Noy N.F., Musen M.A., 2000. "PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment". *In proceeding of the 17th national conference on artificial intelligence (AAAI 2000)*.

Object Management Group (OMG). *"UML semantics, version 1.1."*, September 1997.

Pavel Shvaiko, and Jerôme Euzenat, 2008. "Ten Challenges for Ontology Matching". *In Proceedings of The 7th International Conference on Ontologies*, DataBases, and Applications of Semantics (ODBASE).

Predoiu, L., Feier, C., Scharffe, F., de Bruijn, J., Martin-Recuerda, F., Manov, D., Ehrig, M., 2006. *"State-of-the-art Survey on Ontology Merging and Aligning V2."* EU-IST Integrated Project (IP) IST-2003-506826 SEKT: Semantically Enabled Knowledge Technologies, University of Innsbruck.

Rahm E. and Bershtein P., 2001. *"A survey of approaches to automatic schema matching"*. VLDB journal, the international journal on very large data bases.

Rumbaugh J., Ivar Jacobson, Grady Booch, 1998. *"The Unified Modeling Language"*, Reference Manual. Addison-Wesley.

Ruth Breu, Radu Grosu, Franz Huber, Bernhard Rumpe, Wolfgang Schwerin, 1997. *"Towards a precise semantics for object-oriented modeling techniques"*. In Proceedings ECOOP'97 Workshop on Precise Semantics for ObjectOriented Modeling Techniques, pages 53–59.

Sana Châabane, Wassim Jaziri, Faïez Gargouri, 2013. "A formal framework for modeling topological relations of spatial ontologies", *The Second International Conference on Intelligent Systems and Applications*: INTELLI 2013, Venice, Italy.

Wei Xu, Sisi Zlatanova, Arta Dilo, Peter van Oosterom, 2008. "Modelling emergency response processes: Comparative study on OWL and UML", *Proceedings of the Joint ISCRAM-CHINA and GI4DM Conference*, Harbin, China.