Published in: AGILE 2009 Pre-Conference Workshop
"Challenges in Geospatial Data Harmonisation",
2nd June 2009, Hannover, Germany

Harmonisation of Spatial Semantic Integrity Constraints

Stephan Mäs

AGIS - Arbeitsgemeinschaft GIS, University of the Bundeswehr Munich,
Werner Heisenberg Weg 39, 85577 Neubiberg, Germany
{Stephan.Maes}@unibw.de

Introduction

As part of the quality description of a spatial data set logical consistency is very important to support interoperability between different systems. Logical consistency specifies "the degree of adherence to logical rules of data structure, attribution and relationships" [1]. For the evaluation of logical consistency of spatial object related data integrity constraints play a major role. As part of the data model they assure that the data conform to the structure and the semantics intended by the model. Integrity constraints can be categorised according to the conditions they specify [2][3]:

- **domain constraints** restrict the allowed types of values of an attribute. Spatial information requires data types and corresponding domain constraints which restrict the defined geometric and topological primitives, e.g. polygons or line strings.
- **key and relationship constraints** refer to the possibility to define key values (i.e. unique values) for entity classes, cardinalities for relationships between entity classes and participation requirements.
- **semantic integrity constraints** are explicitly specified and usually more complex. They refer to semantics of the modelled entity classes, which are not representable through the other two categories. They specify relations between the modelled concepts which are usually not explicitly represented in the data. Spatial semantic integrity constraints specify topological, metric, directional or shape restrictions between the involved entities or on specific properties of a single entity. A detailed analysis of the different kinds of semantic integrity constraints can be found in [3].

Constraints of the former two categories are mostly inherently or implicitly incorporated in the schema of the database. They are also formalised and transferred, if data is published via the standardised OGC Web Feature Service Interface, which is making use of an XML Schema document to describe the transfer schema. The mapping of such constraints during schema transformation is an interesting subject of research (e.g. [4]), but not further investigated here.

A formal description of constraints of the latter category is more complex and has for example been researched in [5][6][7]. This work also requires a formal definition of semantic integrity constraints but the purpose is to compare the constraints to discover conflicts and redundancies and not a transfer of the formalised constraints.

Sample use case

Imagine the following use case: a user wants to integrate two spatial data sets from different vendors for a specific application. Both data sets have corresponding entity classes but their quality evaluations were based on different sets of semantic integrity constraints. Figure 1 shows entity relationship diagrams of the two data sets. The semantic integrity constraints refer to the topological relations between areal entities defined in [8].

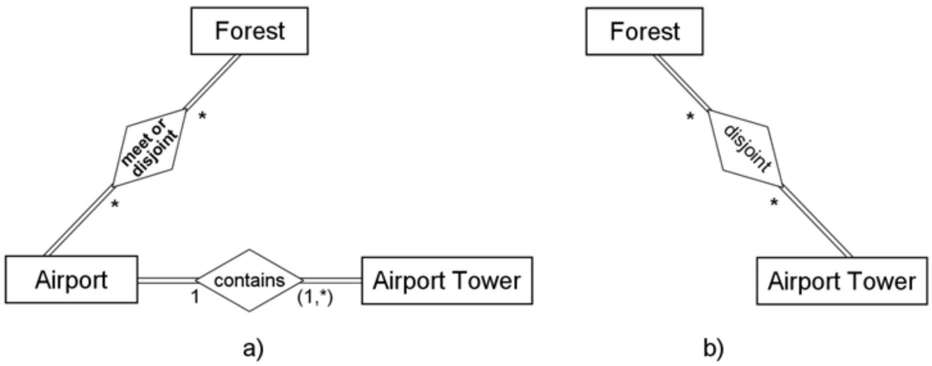

Fig. 1. Entity-Relationship Diagram of the entity classes of two data sets and their topological semantic integrity constraints

Data set a) contains the three classes *airport*, *forest* and *airport tower*. Two pairs of these classes are constrained by topological integrity constraints:

- *Airports* and *forests* are either disjoint or meet.
- Every *airport* contains at least one *airport tower*. Every *airport tower* is contained by an *airport*.

The second data set b) contains only the entity classes forest and airport tower which are constrained by:

- *Forests* and *airport towers* are disjoint.

Before the integration of these two data sets it should be proven, if the sets of integrity constraints correspond to each other with regard to the common classes *forest* and *airport tower*. If the two sets put differing restrictions on the topological relations between the entities of the two classes, the two data models have different semantics, although the same class names are used. In this case it has to be checked if the semantic difference is crucial for the application of the user. If it is, the less restricting data set has to be proven against the corresponding semantic integrity constraints of the other set to reach homogeneous semantic restrictions on both sides. In extreme cases the two sets have conflicting semantic integrity constraints and an integration of the two data sets is not feasible.

For the example shown in figure 1 it is relatively obvious that data sets have the same restriction on the relation between *forests* and *airport towers.* Since in data set a) every *forest* meets or is disjoint from *every airport* the entities of these classes can have intersecting boundaries, but the interiors are disjoint. Since every *airport tower* is contained by an *airport* an airport tower has an intersecting interior with this airport, but no intersecting boundary. Therewith is no intersection between any airport tower and any forest possible, even if this semantic integrity constraint of data set b) is not directly proven in data set a).

The example illustrates that data integration requires some, preferably automatic methods for the comparison of the semantic integrity constraint sets. Depending on the applied spatial relations and the cardinality restrictions, the logical conclusion might become very difficult for human reasoning. If more than three entity classes are involved a manual check is hardly possible. The following chapter gives an overview of a logical reasoning approach, which enables such automatic conclusions. For the lack of space it can only be sketched here; for a deeper insight it is recommended to read [9] and [10].

Reasoning on Spatial Semantic Integrity Constraints

Integrity constraints are defined at the level of entity classes since they always restrict entire classes or subsets of classes. Semantic integrity constraints define cardinality restrictions for a certain, possibly spatial relation between all instances of the involved classes. Thus a formalised description of a semantic integrity constraint must be linked to the instance relations the quality checking procedure applies (e.g. the topological relations in figure 1) and the corresponding cardinality restrictions. A first approach which defined such so called class relations has been made in [11]. As shown in [9] and [12] such class relations can be used to formally define semantic integrity constraints.

In [9] and [10] a reasoning approach to detect conflicts and redundancies in sets of such class relations has been published. The basic idea of this approach is to analyse the network of binary class relations in analogy to constraint satisfaction problems in networks of instance relations (e.g. [13]). Thereby the instance relation and the cardinality restrictions of the class relations are independently analysed regarding logical properties like symmetry (e.g. if A contains B, then B is inside of A) and composition (e.g. if A meets B and B contains C than A is disjoint from C). For the cardinality restrictions these logical properties are derived in [10]. Through this independency the overall approach can be applied with different instance relations and is therewith useful for many spatial semantic integrity constraints.

Contributions to the Workshop

This work comprises the topics of conceptual schema mapping, harmonisation of quality evaluation processes, semantic interoperability and formal semantics, which have been listed in the call for papers of the workshop. The mentioned class relation reasoning concepts will be briefly sketched with some examples, but the main focus of the presentation will be on their application, when sets of semantic integrity constraints are compared and harmonised. The aim is to show that this comparison is absolutely necessary, when object related data from different sources and different quality requirements is integrated. It can also serve to discover semantic distinctions and conflicts between the compared data models. The practical applicability of the reasoning approach can be demonstrated with an implemented prototype at the workshop.

References

1. ISO 19113, Geographic Information – Quality Principles
2. Elmasri, R., Navathe, S. B., "Fundamentals of database systems", 2nd Edition (Addison-Wesley), 1994, pp. 638-643.
3. Mäs, S.; Reinhardt, W.: Categories of Geospatial and Temporal Integrity Constraints. In: International Conference on Advanced Geographic Information Systems & Web Services GEOWS '09, pp.146-151, IEEE computer society
4. Staub, P; Rudolf Gnägi, H.R.; Morf, A: Semantic Interoperability through the Definition of Conceptual Model Transformations. In: Transactions in GIS, vol.12-2, pp.193-207, 2008
5. Louwsma, J., Zlatanova, S., Lammeren, R.v., Oosterom, P.v., "Specifications and implementations of constraints in GIS- with Examples form a Geo-Virtual Reality System", GeoInformatica, 10, 2006, 531-550.
6. Mäs, St.; Wang, F.; Reinhardt, W.: "Using Ontologies for Integrity Constraint Definition". In: 4th International Symposium On Spatial Data Quality, ISSDQ, Peking, China – Proceedings, pp 304-313; August 2005.
7. Wang, F., Reinhardt, W., "Extending Geographic Data Modeling by Adopting Constraint Decision Table to Specify Spatial Integrity Constraints", AGILE 2007, Lecture Notes in Geoinformation and Cartography.
8. Egenhofer, M.; Herring, J.: Categorizing Binary Topological Relationships Between Regions, Lines, and Points in Geographic Databases. Technical Report, Department of Surveying Engineering, University of Maine, Orono, ME. (1991)
9. Mäs, S.: Reasoning on Spatial Semantic Integrity Constraints. In: Winter, S., Duckham, M., Kulik, L., Kuipers, B. (eds.) COSIT 2007. LNCS, vol. 4736, pp. 285–302. Springer, Heidelberg (2007)
10. Mäs, S.: Reasoning on Spatial Relations between Entity Classes. In: 5th International Conference Geographic Information Science (GIScience), Park City, Utah, USA; Lecture Notes in Computer Science 5266, T.J. Cova et al. (Eds.), pp. 234–248. (2008)
11. Donnelly, M., Bittner, T.: Spatial Relations Between Classes of Individuals. In: Cohn, A.G., Mark, D.M. (eds.) COSIT 2005. LNCS, vol. 3693, pp. 182–199. Springer, Heidelberg (2005)
12. Tarquini, F.; Clementini, E.: Spatial Relations between Classes as Integrity Constraints. In: Transactions in GIS, Volume 12 Issue s1, Pages 45 – 57, 2008
13. Egenhofer, M.: Deriving the Composition of Binary Topological Relations. In: Journal of Visual Languages and Computing 5 (2): 133-149, 1994.