Computers & Geosciences 103 (2017) 122–132

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

Rule-based topology system for spatial databases to validate complex geographic datasets

J. Martinez-Llario[a,*], E. Coll[a], M. Núñez-Andrés[b], C. Femenia-Ribera[a]

[a] Department of Cartographic Engineering, Geodesy and Photogrammetry, Universitat Politècnica de València, Camino de Vera s/n, Valencia 46022, Spain
[b] Departamento de ingeniería del terreno, cartográfica y geofísica, Universitat Politécnica de Catalunya, Barcelona, Spain

ARTICLE INFO

Keywords:
Topology
Spatial Databases
GIS
Cartography
INSPIRE

ABSTRACT

A rule-based topology software system providing a highly flexible and fast procedure to enforce integrity in spatial relationships among datasets is presented. This improved topology rule system is built over the spatial extension Jaspa. Both projects are open source, freely available software developed by the corresponding author of this paper.

Currently, there is no spatial DBMS that implements a rule-based topology engine (considering that the topology rules are designed and performed in the spatial backend). If the topology rules are applied in the frontend (as in many GIS desktop programs), ArcGIS is the most advanced solution. The system presented in this paper has several major advantages over the ArcGIS approach: it can be extended with new topology rules, it has a much wider set of rules, and it can mix feature attributes with topology rules as filters. In addition, the topology rule system can work with various DBMSs, including PostgreSQL, H2 or Oracle, and the logic is performed in the spatial backend.

The proposed topology system allows users to check the complex spatial relationships among features (from one or several spatial layers) that require some complex cartographic datasets, such as the data specifications proposed by INSPIRE in Europe and the Land Administration Domain Model (LADM) for Cadastral data.

1. Introduction

In recent years, certain regulations, including the INSPIRE data specification (INSPIRE, 2010) and its subsequent regulations, or the Land Administration Domain Model (LADM) – ISO 19152 (ISO, 2012) are forcing GIS datasets to be highly structured to consider the spatial relationships among features (geographic objects).

Before structuring the datasets using these models, it is necessary to perform a quality control check to assure the spatial relationships. One of the ways to accomplish this task is to use topology rules to model all of the spatial relationships. We will then be able to test and ultimately edit and fix the geometries to build these complex datasets in a non-traumatic way.

Topology is a branch of mathematics that studies the relative positions among geometric features, particularly focusing on the spatial relationships that are maintained when the embedding space is altered with topological transformations, such as translation, rotation and scaling. The GIS community has long understood that topology can help manage spatial data; hence, how to efficiently apply its principles has been a largely discussed issue in the GIS world and is not yet definitely solved.

In this section, we overview its main topics. The second section describes the different topology software approaches. The created topology rule system is presented in the third section. Section four studies a real cartographic model case.

1.1. GIS basic topological structures

Vector-based GIS models spatial features from the real world by means of geometry primitives such as points, lines and polygons. To accomplish this, several abstractions or simplifications must be performed, and consequently a series of conventional rules must be satisfied to coherently characterize the geographic reality. Standardization has been already achieved with OGC Simple feature access (SFA)—Part 1: Common architecture (OGC, 2011), providing a widely adapted way to describe geometry (Yan et al., 2011). It repeatedly relies on topological concepts to define geometry representation, but it not addresses how to store topologically structured data. It was published as the ISO 19125-1 Standard (ISO, 2004). Each geometry is independently described by a set of coordinates

* Corresponding author.
E-mail addresses: jomarlla@cgf.upv.es (J. Martinez-Llario), ecoll@cgf.upv.es (E. Coll), m.amparo.nunez@upc.edu (M. Núñez-Andrés), cfemenia@cgf.upv.es (C. Femenia-Ribera).

http://dx.doi.org/10.1016/j.cageo.2017.03.013
Received 7 October 2015; Received in revised form 19 December 2016; Accepted 14 March 2017
Available online 18 March 2017
0098-3004/

(Worboys and Bofakos, 1993) and sharing geometries between features (geographic objects) is not precluded.

Conversely, the arc-node model is the paradigmatic topological structure for GIS topology. In this model, the primitives are nodes, arcs and faces, all of which are interrelated. As a result, in the arc-node model, relationships such as connectivity, direction, and adjacency are explicitly stored (Murray, 2009). This is a very old model, but the still recent ISO 13249-3 or SQL/MM (ISO, 2016) represents its evolution. The SQL/MM standard defines a way of modelling and accessing a topological dataset and allows multiple levels of topological connectivity.

Both models (ISO 19125-1 or ISO 13249-3) present advantages and disadvantages, so the usage of each model depends on the task at hand. The main disadvantage of ISO 19125-1 is the repeated storage of features, but its structure is less complex and can be seamlessly used for display tasks. The main advantages of explicit storage of the topology structure model (ISO 13249-3) are the avoidance of redundant data storage and the ability to maintain data consistency by performing validation rules (De Hoop and van Oosterom, 1992).

The standard ISO 19125-1 is also known as the OGC Simple Feature Access (OGC-SFA) (OGC, 2011) Model. In addition, the standard OGC-SFA has a more recent version than ISO 19125-1.

1.2. Topology rules

A rule-based topology model consists of a set of spatial rules that constrain the spatial relationships among the features. The GIS user is able to choose which relationships are relevant for the data. After validating the rules, the GIS user handles the cases where the rules are violated and fixes the errors with the aid of topological editing tools. When a geometry that participates in some rule is added, deleted or updated, the spatial extent affected is flagged as a “dirty area” to note that this area should checked again (ESRI, 2003).

With a rule-based topology engine, it is possible to model the spatial relationships among datasets easily. A few real cases of some topology rules area as follows:

- Buildings (polygons) from a dataset must be inside cadastral parcels (polygons)
- Lakes (polygons) and land parcels (polygons) from two different layers must not overlap
- Vegetation (polygons) and soils (polygons) must cover each other
- Endpoints of electric lines (lines) must be capped by transformers (points)

A rule-based topology model can be built over both basic GIS structures: a simple feature model (ISO 19125-1) or an explicit topological model (ISO 13249-3).

If the rule-based topological model is built over an explicit topological model, then some spatial relationships are automatically maintained, and there is no need to design some topology rules: the detection of closed polygons, overlapped areas, dangles, undershoots, points inside polygons (Baars et al., 2004).

2. Topology software background

GIS software packages have worked with topology differently. In the present section, we offer a brief overview of how the theoretical aspects of planar spatial data topology have been transferred to GIS software. Table 1 shows the main characteristics related to the geometry structure and the rule-based topology system used by the most important GIS desktop and spatial databases.

To be able to justly compare all software, we consider that the ArcNode or SQL/MM systems implement some topology rules in an inherent way, although the number of topology rules is much lower (few or very few) than in a full rule-based topology system.

The rule-based topology system column has the following meaning:

- The software labelled “very basic and basic” shows that although they do not implement a pure rule-based topology system, they have an explicit topology system that allows certain SQL attribute queries to be built to check spatial relationships between features as stated before. If the software follows the SQL/MM standard, these SQL queries can be extended to use some relationships between layers because SQL/MM allows multiple levels of topological connectivity.
- The label “medium” shows that even if the software implements some type of topology rule system, the number of topology rules is restricted, and a dirty area system is especially not supported; therefore, every time a topology rule is validated, all features must be checked, which makes the system very inoperative.
- The only two software programs that support a full rule-based topology system (Section 3.5) are labelled as “advanced”.

2.1. Desktop GIS

As mentioned before, ArcInfo coverages (arc-node model) maintain topologically structured data. However, some disadvantages including slowness when assembling features on the fly and constant validation after editing have pushed ESRI to implement another approach to topology to be applied on a simple feature model.

ArcGIS follows a rule-based topology model in which topology is handled via a set of rules that are applied to the feature classes that form the geodatabase.

The open GIS world has in GRASS[1] its major counterpart. GRASS has usually defined itself as a topological GIS (Neteler et al., 2012). Its native format is based on arc-node representation. When importing spatial data stored in a format without topology, it is able to automatically build topology, although it also allows work with spaghetti data if the size of the dataset affects the system.

GvSIG[2] and Kosmo[3] have both implemented topology extensions with graphical user interfaces that provide a range of functionalities to clean and validate data. Like ArcGIS, they are based on a rule-based topology model. In both cases the JTS library (Erickson and County, 2009) provides the necessary algorithms to perform the basic geometry calculations.

QGIS[4] follows the same approach as gvSIG or Kosmo, but the geometry library is not JTS based (Java Topology Suite) but OGR (OpenGIS simple features Reference implementation).

2.2. Moving on to DBMS

As shown before, GIS vendors have conventionally lead the research of topology implementation for spatial data. Nevertheless, van Oosterom et al. (2002) argued why topology management should be carried out within the database: topology is a general and reusable aspect of the data, and thus it should be maintained at the database layer. Different reasons encourage the usage of databases instead of files systems: transactional integrity, multiple users, unified storage and solid SQL standards are the most prominent advantages.

Owing to the spread trend of maintaining spatial data in RDBMS, these products have started to support topology. Oracle Spatial 10 g[5] pioneered a two-dimensional topological architecture in a DBMS environment, supporting primitive topologies including nodes, arcs and faces stored in persistent topology tables.

Radius topology developed by 1Spatial[6] extends the Oracle Spatial

[1] http://grass.osgeo.org/
[2] http://www.gvsig.com
[3] http://www.opengis.es/
[4] http://www.qgis.org
[5] http://www.oracle.com
[6] http://1spatial.com/