

The filling of chosen layers of topographical database by generalization from geospatial databases of higher detail

*Ing. et Mgr. Otakar Čerba, Ing. Karel Jedlička
Department of Mathematics, Geomatics section
Faculty of Applied Sciences
University of West Bohemia in Pilsen
Univerzitní 22
306 14 Pilsen
Czech Republic
E – mail : ota.cerba@seznam.cz, smrcek@kma.zcu.cz*

Abstract

Geospatial data models in general are source bases of cartographic outputs these days.

There is a trend of building robust fundamental data models in geospatial data modeling. These models are sources for building data models with lower detail.

The article is focused to possibility of using digital cadastral map as a source of geospatial data for chosen classes of objects of topographical database. The text describes selection of suitable techniques of model generalization of point, line and polygon data representation from source database. The aim of the text is to choose techniques, which allow automatization of essential part of the whole technological process.

Abstrakt

Geoprostorové datové modely obecně jsou v dnešní době zdrojovými bázemi pro generování kartografických výstupů.

V oblasti geoprostorového datového modelování je patrný trend budování robustních fundamentálních datových modelů ze kterých se následně odvozením vytváří datové modely menší podrobnosti.

Článek se konkrétně zabývá možností využití katastrální mapy v digitální podobě jako zdroje geoprostorových dat pro vybrané třídy objektů topografické databáze.

Jeho obsahem je výběr vhodných technik pro modelovou generalizaci bodových, liniových i polygonových reprezentací ze zdrojové databáze. Účelem textu je vybrat takové techniky, které budou umožňovat co největší automatizaci celého technologického procesu.

Introduction

Our team is now co-working in a project, which is focused to build an alternative parcel data model for real estate register of Czech Republic. There is an idea to build this model in a way, which would allow using the real estate model for creating and updating other geospatial data models. This paper is concretely focused to build a part of methodology of creating a data model for a topographic map.

There are two fundamental bases of geographical information in Czech Republic now. First is the real estate database (The original abbreviation is DKM); the second is a Fundamental Base of Geographic Data (The original abbreviation is ZABAGED). Both of them are created, updated and administered by Czech Office for Surveying, Mapping and Cadastre (COSMC). Using the first one for filling chosen layers of ZABAGED would have following advantages:

- Possibility to control content of ZABAGED, because of using multiple data sources (i.e. orthophoto together with DKM).
- Updating of ZABAGED based on real changes which are recorded in real estate database.
- Multiple use of real estate database data.
- Probably cheaper way, how to acquire data for ZABAGED.

The following text is about problems in generalization of object representation, which come in generalization from the scale of DKM to the scale of ZABAGED. The text also considers different thematic content of those databases. The paper shortly describes theoretical background of generalization types and characters and proposes concrete algorithms and workflows to use in DKM to ZABAGED generalization. The core of the text is about geospatial generalization. The cartographical part of generalization is not a main subject of the article.

We are not telling that it is possible to create a topographic database (or map) just from a cadastral database. But it can be a large source of geographic data for ZABAGED layers. We are also far to say that it is possible to create chosen ZABAGED layers in a full automatic way, but there definitely are processes, which can be automated. Those processes can create a raw geospatial data model. See next chapters for detailed information.

Generalization types and characters

Generalization is a process, which allows creating a less detailed geospatial data from higher detailed data.

Geospatial database is a digitally stored structure of attribute and geospatial data.

Topographical database means geospatial database with a topographic context.

The difficulty of creating less detailed map from a map of higher detail depends also on dominant character of generalization. We distinguish between two characters of generalization, qualitative and quantitative.

We talk about the **qualitative character of generalization** when there is a change of object representation (like from polygon to line, polygon to point, etc.). The selection also reduces a big amount of source data. We can also say that generalization has qualitative character when the map purpose is changing¹.

Quantitative character of generalization means that just the flat system of points, lines and polygons is simplified. The map purpose remains the same, like topographic map remains topographic, just in less detailed scale.

We talk about the **character** of generalization because there is no strict distinction between qualitative or quantitative generalization².

We also can divide generalization to topographic or cartographic.

Topographic generalization is applying when we classify reality and select its representation in a map (or data model) during a topographic mapping.

Cartographic generalization is a generalization between cartographic models (maps) of different scales.

The division of generalization is displayed in the figure 1. There we can see that generalization from cadastral data model to topographic map is cartographic and has qualitative character.

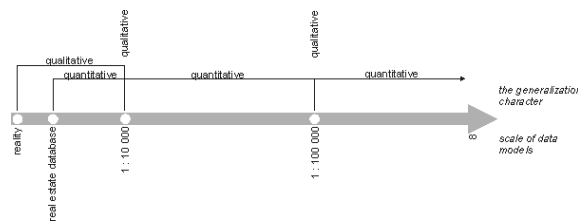


Figure 1 – The division and character of generalization.

We use new term **geospatial generalization**. This means a process which runs between two models of geospatial databases (usually of different scales) and which has a **geospatial model** as its result.

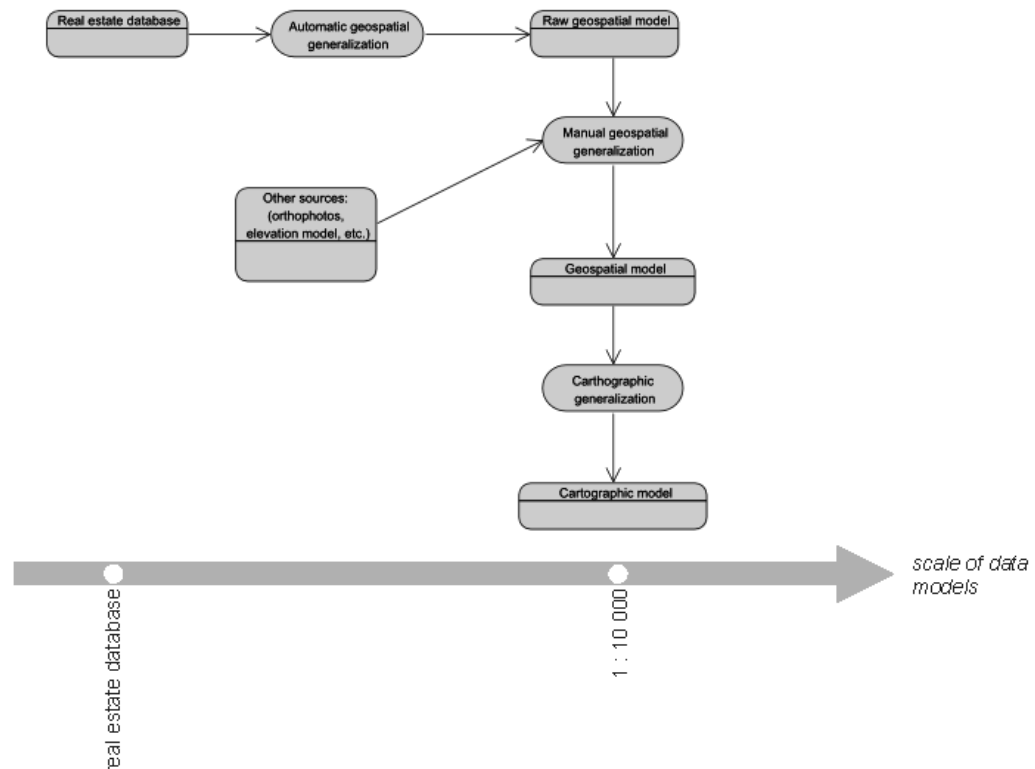


Figure 2 – The workflow of generalization process from cadastral to topographic data³.

Geospatial model is a product of geospatial generalization and it's a model which doesn't deform geometric position of its features more than is necessary because of the model scale. The geospatial generalization just simplifies the shape of features from source model to be portray-able in the resultant model. But it doesn't solve any geometrical or topological conflicts among features. We can also say that every point, line vertex or polygon vertex, which remains in resultant geospatial data model, has the same position as it had in the source geospatial data model.

In contrast to geospatial model is **cartographic model**, which is a product of cartographic generalization, has solved conflict among features. So it's possible to directly print a map from that model.

We find a difference between geospatial and cartographic generalization when we take a closer look to figure 2. The first one produces a geospatial model with accurate coordinates of each its element (in sense of point and line or polygon vertex). The next cartographic generalization runs between models of the same scale and its focused to solve conflicts of features by displacing them, and to other cartographical procedures like label placement and others. Those are operations which are hard to implement as automatic processes in computer and there is still necessary an interactive work of educated cartographer to get a correct result. So we can say that geospatial generalization is a preprocessing for cartographic one.

We divide geospatial generalization into two parts, automated and manual. The goal of the first one is to create a raw geospatial model without an user intervention. That type of model will necessarily content errors, because of errors in source data which you cannot avoid in reality. That's the reason why the interactive user work is necessary during creation of the final geospatial model⁴. The user checks the raw model, i.e. comparing with the orthophoto or new mapping in selected areas, and he corrects the model interactively. There can also enter other features to the final geospatial model then those which are in the real estate database. The contour lines and other elevation features can be a good example.

Generalization techniques

We distinguish following base techniques which we can use in geospatial generalization:

- selection,
- classification,
- geometric generalization,
- collapsing,
- operations with polygons,
- attribute generalization.

These techniques changes the data model, but no more than is necessary because of scale change. The other generalization techniques are making bigger changes in data model and we assign them to the cartographic generalization process because of that.

We can do the **selection** in the normative way or based on minimal portray-able limits. The minimal limits could be quantitative (length, perimeter, area, distance from other features, etc.) or qualitative (importance, theme, etc.). The normative way is based on experimentally acquired standards, which determine how many features of each theme (or

class) could be displayed in the map (model). In our particular case, we can use just the selection based on minimal limits, because the influence of exaggeration is not significant in the scale of the resultant topographic model (1:10 000).

We can apply the selection to all three basic components of map – points, lines and polygons. There exist special selection techniques for concrete features (such as rivers or roads) to keep their characteristic.

The change of **classification** is a specific technique for geospatial generalization. It can merge or split groups (classes) of features.

Geometric generalization consists in adapting shapes of lines or outlines of polygons. We can talk about following geometric operations:

- feature shift,
- rotation,
- shape improvement,
- simplification,
- smoothing.

Sometimes we can see *feature shift* and *rotation* in this generalization category, but we assign them to the cartographic generalization processes in our case (as a map harmonization processes). The reason is that those processes degrade the data accuracy. We can use them in geospatial generalization only in case of correcting source data by comparing with other data source (like real estate data model and orthophoto). The shape improvement we also assign to a cartographic generalization (except right-angled adjustment and substitution polylines by curve).

The *right-angled adjustment* is the next technique, which we can assign to geometric generalization. We can apply the process to generalization of buildings, because there is an assumption that buildings are right-angled. There are two methods of the right-angled adjustment:

- measuring of edges lengths and adjustment using the method of least squares,
- polygon axes based technique.

We apply *simplification* in a case of that the concrete feature detail would not be visible in the scale of the resultant model or map which would be created from that model.

The simplification technique should keep the area and relative proportions of the feature in the output model. Also endpoints have to remain. Otherwise, there would be inconsistent topology in the model. The last thing which should be kept in the result is a characteristic shape of features.

There exist lots of algorithms which can be used for line simplification. They used basic geometric principles for reducing line vertexes. There are some of them:

- deleting (or keeping) every n point in the line (where n means an integer value),
- eliminating of points, which are close to each other,
- elimination of points under an angle threshold,
- elimination of points with a small perpendicular distance from a base line,
- Lang algorithm (based on comparing of minimum perpendicular distances),

- Reumann-Witkam algorithm (based on generalized line buffering),
- Visvalingam-Whyat algorithm,
- Douglas-Peucker algorithm.

The first two algorithms are often used as a preprocessing for the next. This reduces the processing time.

The simplification is often used for deleting unwanted detail of buildings, like too small edges. If the minimal portray-able length is so big that any of building's edge or just one is longer than limit, then the polygon collapses (see below) to a point or there is a maximal geometric generalization applied. This means that the polygon is replaced by rectangle with following properties:

- the rectangle have the same direction as the longest polygon edge,
- the area of the rectangle and of the polygon should be the same,
- the centroid should have the same position,
- the ratio of sides of output rectangle should be the same as the ratio of sizes of rectangle which is circumscribed to source polygon.

The *smoothing* is used in geospatial generalization just, if it is for get better quality line at the output, like a curve instead of a polyline.

The **collapsing** means a reduction of feature dimension. There are three types of collapsing:

- polygon to line (rivers too small to be portray-able as polygon),
- polygon to point (buildings too small to be portray-able as polygon),
- line to point (too short bridge to be portray-able as line).

We mean merging, splitting and deleting polygons, when we talk about **operations with polygons**. There is necessary to modify shape of adjacent polygons after one of those operations ran.

Merging (aggregation) could be applied to features with the same or similar attributes⁵. The inner boundaries are dissolved and the group of features is replaced by one. The process could be based on buffering or algorithms which use triangularization.

The polygon *deleting* can be realized in two base ways:

- counting of center of gravity of the deleted polygon and connecting adjacent polygons boundaries into this center. This is easy to implement way, but it doesn't work well for bigger or concave polygons.
- Algorithms based on triangularization and skeleton finding.

The next basic generalization technique is the **symbolization** change, but it's also assigned to cartographic generalization. So we don't describe it further here.

The last technique, which we haven't been talking about yet, is generalization of text labels. But we would like just to keep all of text elements in the geospatial model and let cartographer with territory knowledge to decide which one to keep and which one to delete.

The feature generalization is composed of particular processes. We would like to build generalization models like these which you can see in figure 3, 4 and 5 for all layers of real estate database which could be usable for ZABAGED creation.

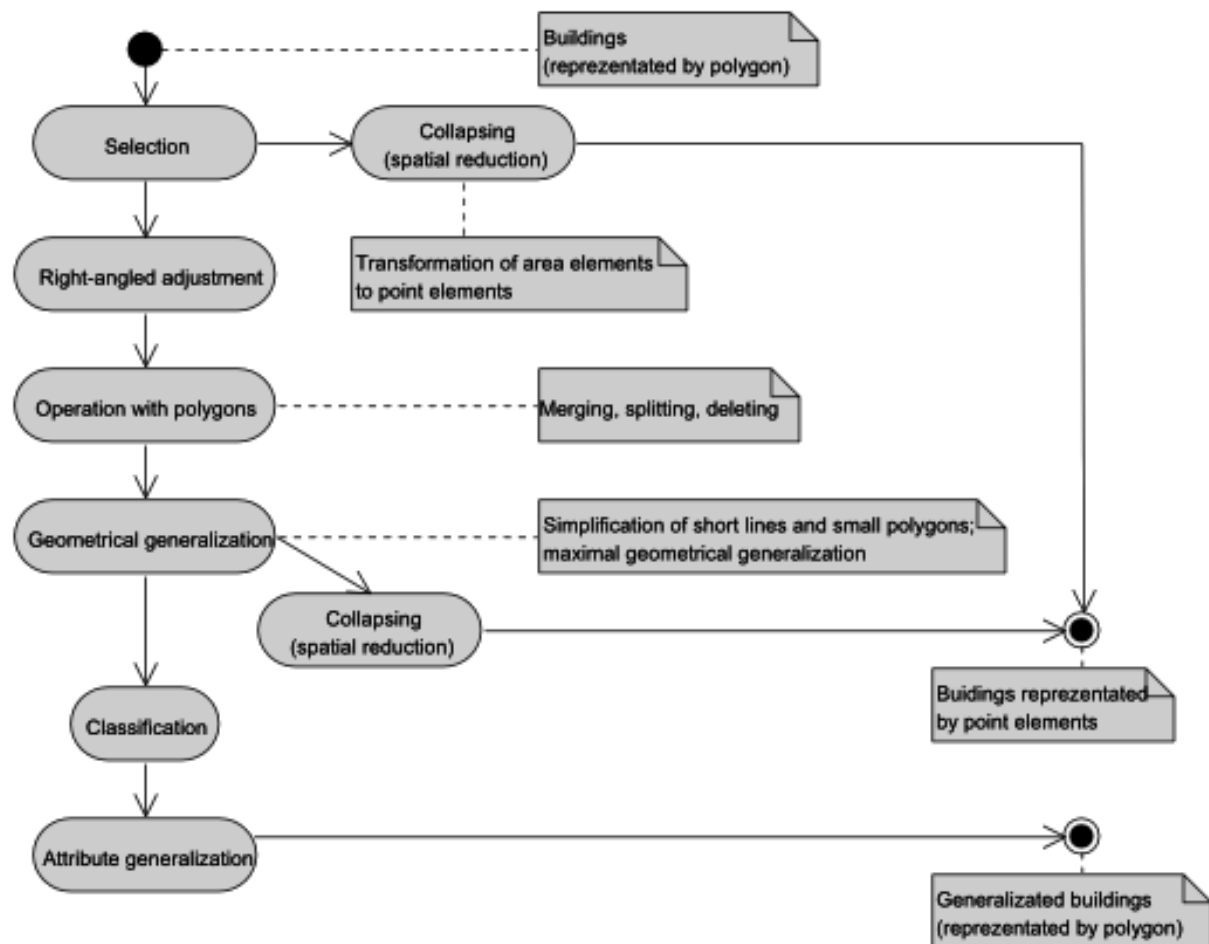


Figure 3 – The process of building generalization.

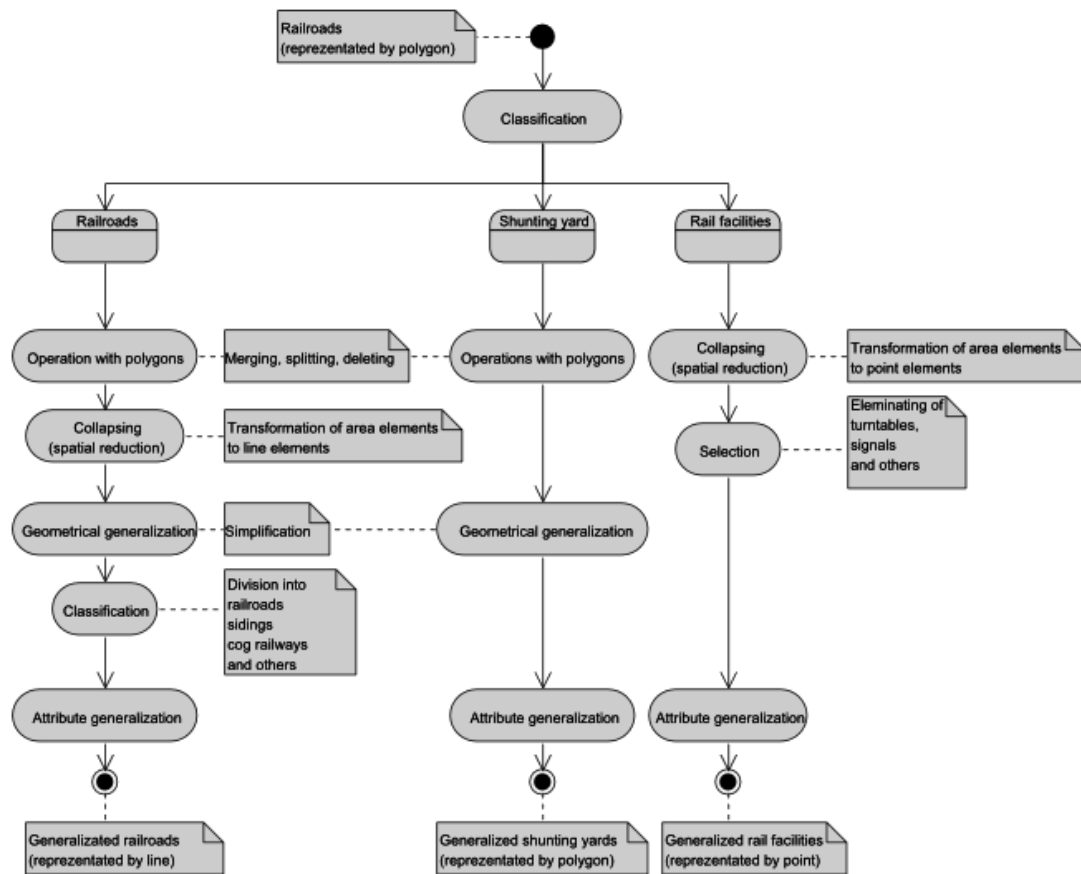


Figure 4 – The process of railroads generalization.

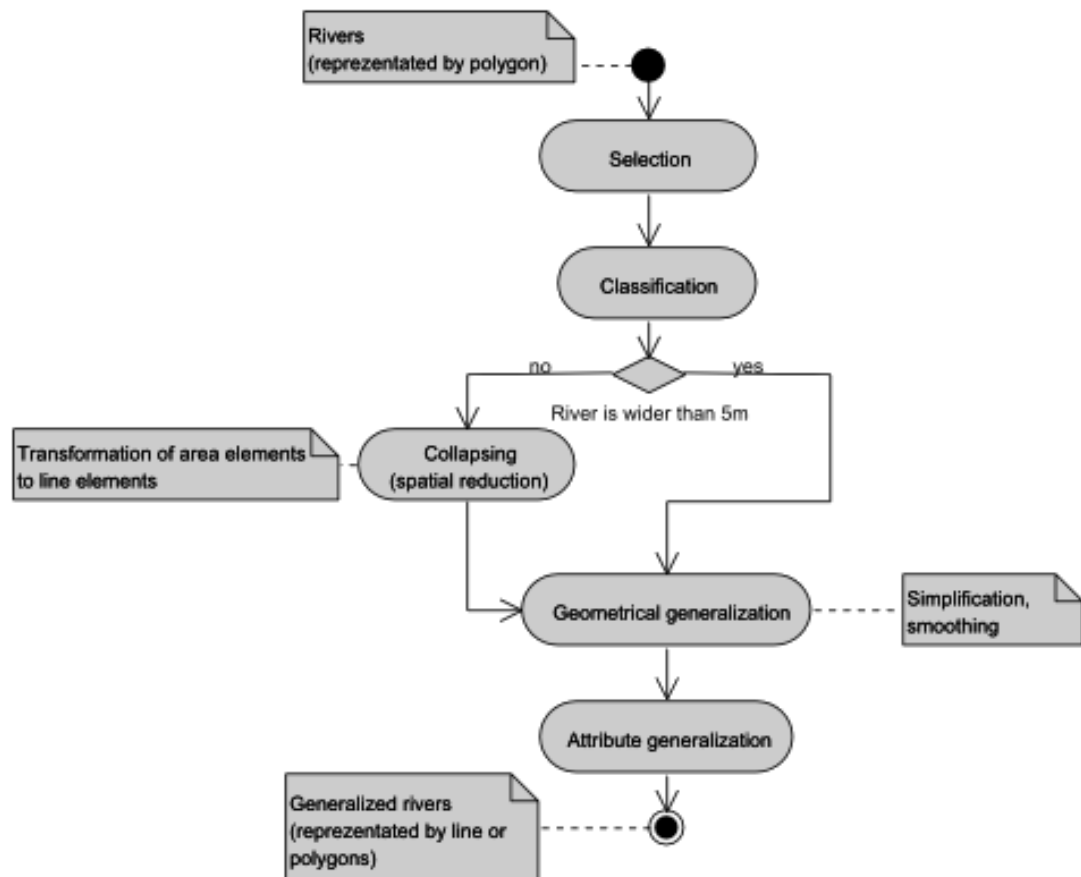


Figure 5 – The process of rivers generalization.

Conclusion

It is easy to implement semiautomatic generalization techniques in object oriented database model. One of the possibilities is to use a XML technology (eXtensible Markup Language). In reality it means to use a combination of XML implementations:

- GML (Geography Markup Language). It is a reference frame for geodata handling. It supports modeling, transportation and storing of geospatial information.
- XSL (eXtensive Stylesheet Language) is especially known because of possibility to transform XML documents to visualizable form. We concern to XSL, respectively to its subset XSLT (eXtensive Stylesheet Language Transformation), because of its ability to transform data between different data models.
- XPath (XML Path Language) allows to search in and reference to XML elements using a tree hierarchy of document.
- DOM (Document Object Model) is one of two base interfaces for work with XML documents, which is represented by classical tree structure where each element is equal to each node.

The transformation principle is really simple. There are geodata as the XML document and transformation style at the entry. The XML form of geodata could be in a universal format, but there is a format adapted for geodata (like the GML) recommended. The description of transformation processes is stored in the XSLT format.

These Entries are processed by the XSLT processor, which generates output document based on conditions from transformation style. The XML technology connected to semiautomatic generalization could be an interesting way to get generalize and import data into ZABAGED.

Even the modern technology can not replace the human work, experiences and knowledge of area of interest. Those are reasons why our raw geospatial model is not ready to use immediately. There remains a lot for manual work which has to be done. There is also necessary to fix errors and inaccuracies which are in input data. But both of our models (raw and final geospatial model) and the ways to build them (automated and manual geospatial generalization) can still be strongly helpful for cartographers and other geographic specialists in their data mining activities.

Used sources

ČADA, Václav, MILDORF, Tomáš. Vymezení základních geoprostorových dat na úrovni pozemkového datového modelu (koncept článku). Internal material of West Bohemian University, 2004 [cit. 7.11.2004].

HARDY, Paul. Active Objects and Dynamic Topology for Spatial Data Re-engineering and Rich Data Modelling [online]. Dagstuhl Seminar 011910 – May 2001 – Computational Cartography and Spatial Modelling, 2001 [cit. 25.10.2004]. http://www.hardy.34sp.com/papers/2001_dagstuhl_pgh.pdf

HAŠEK, Aleš. Soubor topografických map pro služební potřebu. Výzkumná zpráva č. 344. 1st edition. Praha: Výzkumný ústav geodetický, topografický a kartografický, 1969.

HOJOVEC, Vladislav, DANIŠ, Michal, HÁJEK, Milan, VEVERKA, Bohuslav. Kartografie. 1st edition. Praha: Geodetický a kartografický podnik v Praze n.p., 1987. 660 p.

LAUERMANN, Lubomír. Technická kartografie. Díl 1. 1st edition. Brno: Vojenská akademie Antonína Zápotockého, 1975. 346 p.

MONMONIER, Mark. How to Lie with Maps. 2nd edition. Chicago: University of Chicago Press, 1996. 222 p. ISBN 0226534219.

ROBINSON, Arthur H., MORRISON, Joel L., MUEHRCKE, Philip C., KIMERLING, A. Jon, GUPTIL, Stephen C. Elements of Cartography. 6th edition. Wiley, 1995. 688 p. ISBN 0471555797.

VEVERKA, Bohuslav. Topografická a tematická kartografie. 2nd edition. Praha: Vydavatelství ČVUT, 1995. 202 p. ISBN 80-01-01245-X.
Seznam literatury

¹ We understand the term map purpose analogically as (Lauermann). He says that the map scale is dependent of the map purpose. But we can definitely maps with different purpose in one scale.

² We can talk about strictly qualitative or quantitative generalization just in case of particular algorithms. But the whole generalization process is composed from a lot of them and we the generalization character depends on which type is the majority of used processes/algorithms.

³ We can consider real estate database as geospatial data model for our purposes.

⁴ Of course, there is possible to create cartographic model directly from raw geospatial model. The creating and maintaining of clean geospatial data model makes a sense just in case that there exists exploitation of it. This exploitation may not exist in organization, which is focused just to create maps. But a really good exploitation would have model like this for example for facility management systems.

⁵ The classification process could anticipate merging.