SOLAP On-the-Fly Generalization Approach Based on Spatial Hierarchical Structures

Tahar Ziouel, Khalissa Amieur-Derbal, Kamel Boukhalfa

SOLAP On-the-Fly Generalization Approach Based on Spatial Hierarchical Structures

*Tahar Ziouel, **Khalissa Amieur-Derbal, and **Kamel Boukhalfa

*High School of Computer Sciences, Algiers
t_ziouel@esi.dz
**USTHB University, Algiers
{kderbal,kboukhalfa}@usthb.dz

Abstract. On-the-fly generalization, denotes the use of automated generalization techniques in real-time. This process creates a temporary, generalized dataset exclusively for visualization, not for storage or other purposes. This makes the process well suited to highly interactive applications such as online mapping, mobile mapping and SOLAP. BLG tree is a spatial hierarchical structure widely used in cartographic map generalization and particularly in the context of web mapping. However, this structure is insufficient in the context of SOLAP applications, because it is mainly dedicated to the geographic information processing (geometric features), while SOLAP applications manage a very important decision information that is the measure. In this paper, we propose a new structure, SOLAP BLG Tree, adapted to the generalizaion process in the SOLAP context. Our generalization approach is based on this structure and uses the simplification operator. Combining the topological aspect of geographical objects and the decisional aspect (the measure).
Our experiments were performed on a set of vector data related to the phenomenon of road risk.

Keywords: On-the-fly map generalization, Hierarchical spatial structures, Spatial data warehouses, SOLAP

1 Introduction

Business intelligence is a major decision-making tool for strategic and daily management of data in the enterprise. It provides essential information in several forms to users (decision makers) so that they can analyze and manage their business by taking effectives decisions. Data warehousing and On Line Analytical Processing (OLAP) are technologies intended to support business intelligence. Indeed, Analysts and decision makers in the enterprise can thus analyze interactively and iteratively multidimensional data at a detailed or aggregated level of granularity through online Analytical Processing tools, OLAP, [1] [2] [3]. Nevertheless, these data may have a geographic component that OLAP systems cannot process due their lack of tools for managing spatial data. A new technology has so, emerged, Spatial OLAP (SOLAP), resulting from integrating GIS technology (Geographic Information System) and OLAP [4] [5].

SOLAP has been defined by [6] *as a visual platform built especially to support rapid and easy spatio-temporal analysis and exploration of data following a multidimensional approach comprised of aggregation levels available in cartographic displays as well as in tabular and diagram displays.*

SOLAP enriches the analysis of classical OLAP systems capabilities in many ways. For instance, providing visual information through maps and interacting with them by formulating queries directly on the cartographic display. Thus, the cartographic component in OLAP systems represents a graphic interface to spatial data warehouses (SDW) which introduce spatial data as subject or analysis axes.

In this context, the analysis of multi-dimensional spatial data often requires navigating through different levels of detail, in order to study the evolution of a phenomenon (fact), and thus allows an effective decision-making. On-the-fly generalization process is therefore well suited to this context, because it can interactively adapt the visualized geographic information to decision-makers needs [7]. However it only addresses the cartographic aspect, at the expense of the decisional one, which is important in multidimensional analysis.

Widely addressed in cartography, on-the-fly generalization, well suited to highly interactive applications such as SOLAP, consists in generating temporary data at different levels of detail from the most detailed level. Different on-the-fly generalization approaches have been developed [8] and classified in two main groups. The first group relies on fast map generalization algorithms that generate coarser levels-of-detail in real-time. [9]. The second group utilizes hierarchical spatial data structures [10] [11]. To the best of our knowledge, no work has proposed a generalization approach for SOLAP, nor the consideration of the decision making aspect (measure) in the generalization process.

In this paper, we propose an on-the-fly generalization approach for SOLAP systems. The approach we propose integrates topographic appearance (distance) and decision-making aspects (measure) for an on-the-fly generalization suited to cartographic experts and decision makers.

The present paper is organized as follows; the next section introduces some research work related to the addressed issue. Section 3 presents a detailed description of the proposed approach. The different steps of our experiments and some results are described in section 4. Section 5 concludes the paper and presents some perspectives.

2 Related work

Several research work have addressed the generalization for more than three decades [12] [13] [14]. On-the-fly generalization has emerged with the development of highly interactive applications of cartography such as web mapping. The main used operators are selection and simplification [8]. Among research work that have adressed on-the-fly generalization, we can cite [15] [16] carried in the contexte of the European project GiMoDig [17]. The objective of this project is to develop and test methods for providing spatial data to mobile users through real-time generalization. The work presented in [18], combines multiple representation and cartographic generalization and uses an implementation of multi-agent system where each agent was equipped with a genetic patrimony.

Since cartographic generalization creates a hierarchy of levels of detail, it is natural to use hierarchical structures such as tree structures for storage of the geometry (point, line, polygon) of an object in the highest level of detail. This structure is enriched with information that reflect the importance of a hierarchical level, from which, requested levels of detail may be generated. The generalization process is therefore, speeded up with rapid access to the elaborate structures. For each type of spatial data, corresponds an appropriate hierarchical structure that enables interactive and rapid generalization of geographic objects. BLG tree (Binary Line Generalization tree) has been proposed for linear objects [19] [20], it applies the simplication generalization operator that uses a variant of the Douglas-Peucker algorithm [21]; instead of deleting the less important vertices, it stores them in the structure. The GAP tree (Generalized Area partitionning) has been proposed for the selection and fusion of polygons [20] [10] [11]. The Quadtrees have been proposed for point objects point objects, they allow applying the selection, simplification, aggregation and displacement operators [8] [22].

Furthermore, as we have already mentioned in Section 1, and to the best of our knowledge, there are no research work that have addressed integration of generalization in SOLAP. Nevertheless, some work focused on integration of spatial data as dimension or fact in SDW [4] [7].

In this paper, we propose to integrate on-the-fly generalization process in SOLAP, to adapt the level of detail that meets the decision-makers needs. The approach we propose focuses on linear objects that represent rivers, roads, etc. These latters constitute a geographical dimension linked to the phenomenon of road risk that we consider as use case study. BLG tree structure is dedicated to cartographic generalization of linear objects (roads in our case study). However, this structure cannot be efficiently used in decision-making process, because they don't consider the main decisional information in SOLAP, that is *the measure*.

To better understand this problem, we propose the example illustrated in Figure 1. The analyzed map contains six objects with associated measures. As presented in this example, among the objects at the most detailed level , the object C possesses the greatest measure (30), despite its geometric size is not indicative (see figure 1.a). When reducing the scale, the classical generalization process is triggered, considering only the topographic aspect, the object C is imperceptible (see figure 1.b) despite its decision relevance (the greatest measure) compared to the objective of the analysis performed by the decision-makers.

We propose a generalization approach based on a new version of BLG tree adapted to SOLAP called SOLAP BLG tree.

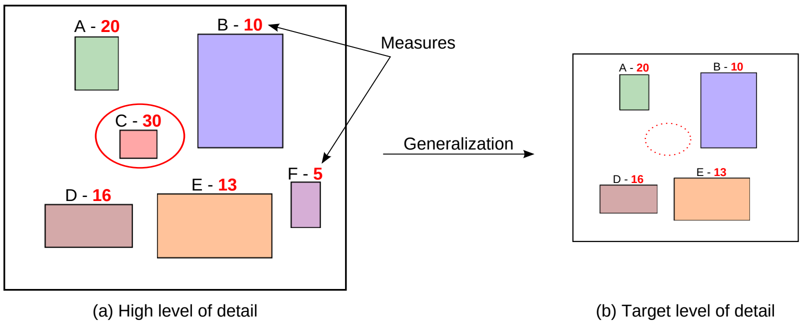

Fig. 1. Traditional generalization results

3 Proposed approach

The main objective of our approach is to develop an on-the-fly generalization system adapted to SOLAP applications. This system must be able to combine the decision and cartographic aspects to produce maps adapted to the needs of decision makers. Figure 2 shows the overall architecture of our approach. The spatial data warehouse stores decision data (measures, fact, dimensions, etc.) and cartographic data. The latter represented in a single level of detail (the highest one). When the user sends his request, the result is extracted from the stored data. It does not necessarily reflect the level of detail requested by the decision maker, therefore an on-the-fly generalization process is necessary to adapt this result to the expressed need.

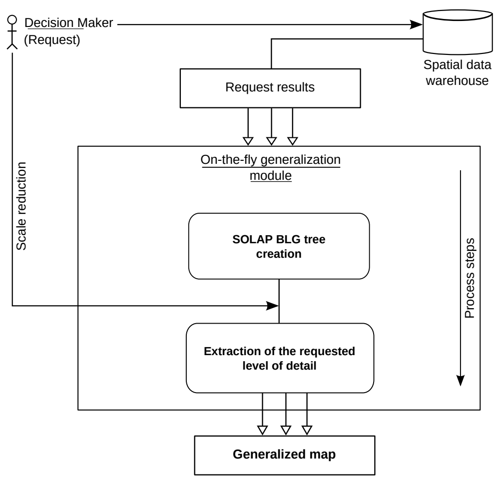

Fig. 2. Global architecture of our approach

The generalization process that we propose is based on SOLAP BLG tree structure. A set of parameters related to the decision aspect is integrated to the structures including the importance function, aggregation etc. All these concepts are described as in the following sections.

3.1 SOLAP BLG tree

The creation of the SOLAP BLG tree revolves around two main steps: (1) the attribution of an importance value to each point of the polyline and (2) Creating the Hierarchy considering the importance of the points. Indeed, a polyline object (road, river, etc.) consists of a set of points (vertices). SOLAP BLG tree stores these points in a hierarchical structure. Each node of the structure consists of a point of the polyline along with its importance value elaborated by the following function: the importance $I(p)$ for each point p will be determined according to its distance $D(p)$ and its associated measure $M(p)$ as follows: $I(p) = f(D(p), M(p))$. This function can be described by the sum of its cartographic importance (distance) and its decisional importance (measure):

$$I(p) = D(p) + M(p) \tag{1}$$

The distance $D(p)$ is the orthogonal distance between the segment connecting the two end points and the point p of the polyline. $M(p)$ represents the measure at the point p. A node in the structure is created to represent a point p_i, which importance value is $M(p_i)$.

The polyline (p_1, p_n) will be processed as follows: If the node root is represented by p_k (a point on the polyline) having the highest importance value, the creation of other nodes follows an iterative process addressing all the points of the segments $[p_1, p_k]$ and $[p_k, p_n]$. To illustrate this process, we propose the following example on road risk analysis; we focus on the number of accidents recorded on road segments connecting ten cities represented by c_1 to c_{10} points (see Figure 3). Each segment carries a measure that represents the number of accidents reported on the segment connecting city c_i to city c_{i+1}.

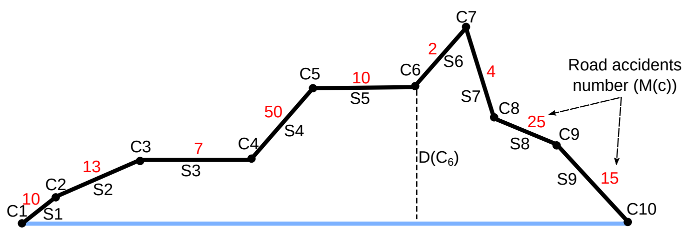

Fig. 3. Polyline (road) representation in SOLAP

We emphasize that in this case, the measures are associated with road segments, or the process requires their transposition to the endpoints constituting these segments. To do this we propose that the measure of each point p_i is determined by the maximum value of the measure of the segments to which the point pi belongs: Let $M(p_i)$ be the measure at point p_i.

$$M(p_i) = Max(M([p_{i-1}, p_i]), M([p_i, p_{i+1}])$$

Furthermore, the values of the measures and distances such as identified have different domains. Indeed there is a significant difference between these two parameters. A normalization step is thus necessary, in order to make the values comparable to each other. To do this we will restrict values between 0 and 1. For each value V of a measure or a distance, its normalized value V' is calculated as follows:

$$V' = \frac{V - V_{min}}{V_{max} - V_{min}} \tag{2}$$

Thus, the BLG structure of the original polyline depicted in figure 3 is as shown in figure 4.

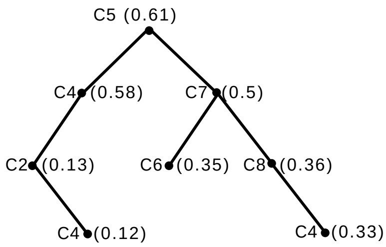

Fig. 4. The built SOLAP BLG tree structure

3.2 Proposed generalization process

The on-the-fly generalization process adapted to SOLAP context is guided by the SOLAP BLG tree structure according to the method described below.

Once these structures built, they are saved in a session work. When navigating between the different levels of detail, the on-the-fly generalization process is triggered to retrieve only the visible points in the required level of detail. The selection of these points is performed by comparing the importance values previously stored with a threshold value. The latter is determined by the visualization scale and other cartographic parameters that require the intervention of an expert cartographer. In the context of this work we used experimentally determined thresholds. Indeed, the threshold determines the tree traversal depth, by selecting only the nodes whose relevance value is greater than the threshold.

Measures aggregation The measures associated with the different objects are subject to an aggregate function that determines the measures of the resulting objects. This maintains the importance of the decision-making aspect of the different requested levels of detail. This aggregation function is developed according to the analyzed fact. For example for the analysis of the road risk phenomenon, the proposed aggregation function is the sum function, to preserve the information on the total number of accidents on the generalized object.

To aggregate polylines measures, we propose the creation of a data structure, containing the values of the measures associated with the different segments constituting the initial polyline depicted in Figure 3 as shown in the example of table 1.

Table 1. Data structure dedicated to measures storage

Segment	S1	S2	S3	S4	S5	S6	S7	S8	S9
Measure	10	13	7	50	10	2	4	25	15

During the generalization with the BLG tree, we obtain a new polyline where the segments S_i to S_k are removed and replaced by a new segment formed by the first point of S_i and the last point of S_k, to evaluate the measure associated with this new segment, one can read the above table and sum the measures corresponding to the segments from S_i to S_k.

To illustrate this process we will use the polyline shown in Figure 5, in each segment we took the values of measures in accordance with the table above. After simplification of the polyline we get two segments formed by the points C_1, C_5 and C_{10}, the points C_2 through $C4$ are deleted along with the points C_6 through C_9. The measure associated with the new segment $[C_1, C_5]$ represents the sum of the measures contained in the table (measure of the segments S_1 through S_4).

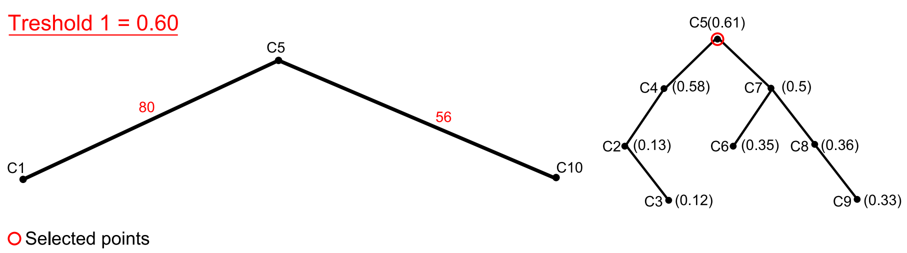

Fig. 5. SOLAP BLG tree generalization results

It is therefore clear that the displayed map will be simplified in order to highlight the information requested by the query.

4 Experimentation

The validation of our approach involves the construction of the proposed spatial structure SOLAP BLG tree used in in the implemented generalization process. We choose the road risk as a case study given its socio-economic impact worldwide. According to statistics from the World Health Organization [23], the road causes each year more than 1.2 million deaths and between 20 to 50 million wounded. In a previous work, we addressed this phenomenon by incorporating spatial information [24] [25].

Our tests are performed on vector spatial data. We used multiple softwares and hardawre ressources to implement our generalization prototype : (1) Oracle 11g Enterprise Edition as DBMS (Management System Database) via its component Oracle Spatial. (2) Oracle MapViewer for viewing the map of the analyzed area. (3) Oracle Weblogic Server on which MapViewers components are deployed. (4) Oracle Map Builder was used to load the geographical data in the DBMS and the construction of the map. (5) Oracle JDeveloper tool as a code editor.

Our experiments were performed on a data set that represents the road theme of Dar El Beida municipality in Algiers enriched by different measures representing the number of accidents recorded on the considered roads (Figure 6).

4.1 SOLAP BLG tree Test

To test the SOLAP BLG tree, we selected a road in Dar El Beida municipality. This road is shown in red in Figure 6. It includes 28 segments each one having a measure. Our generalization system simplifies this road at smaller scales, taking into account the decision aspect (measures).

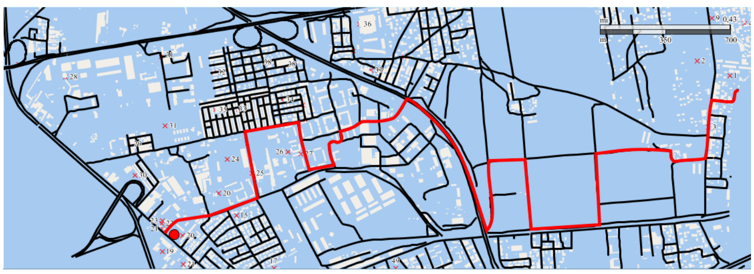

Fig. 6. Road network of Dar El Beida and the selected road

Figure 7 illustrates a detailed representation of the selected route; segments with the highest number of accidents are highlighted.

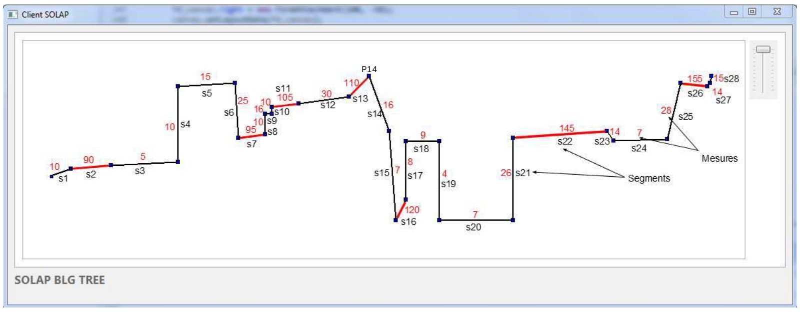

Fig. 7. Detailed description of the selected raod

The SOLAP BLG tree corresponding to the selected route is shown in Figure 8. The root node contains the point p_{14}, which has the highest importance value. Points stored in the top levels are the points having the highest importance values.

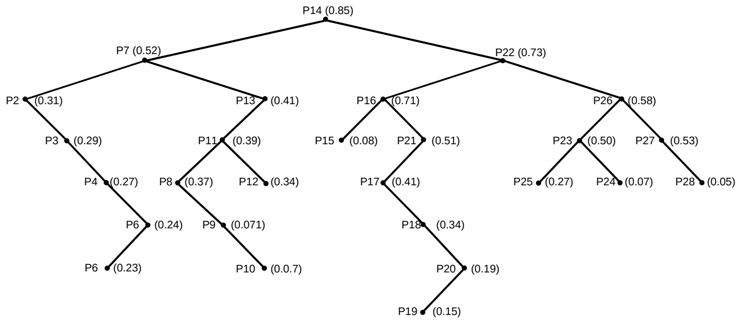

Fig. 8. SOLAP BLG tree corresponding to the selected road

Following the proposed approach, the generalization of the selected road allows to restore only the segments points visible at the required scale. From a more detailed scale, we can analyze the results obtained at different scales (see Figure 5). For example in scale 1: 5000 all relevant segments road except the segment $S2$ are visible on the map, whereas, at the scale 1: 10,000, there are only four relevant segments and at the scale 1: 50,000 relevant segments are no longer visible.

Figure 10 shows a comparison between the results of the generalization with SOLAP BLG as part of this work and the results of the generalization with the classic BLG tree. We can see that at the same scale 1: 10 000, relevant road segments are visible in the case of SOLAP BLG tree (segments S13, S16, S22 and S26), while they are no longer in the case of classical BLG tree, despite

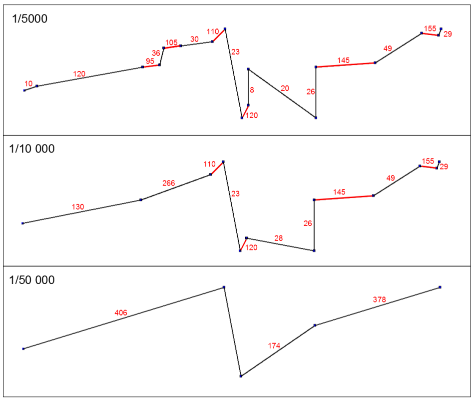

Fig. 9. Generalization results guided by SOLAP BLG tree

their decision relevance, hence the importance of generalization with SOLAP BLG tree in the SOLAP context.

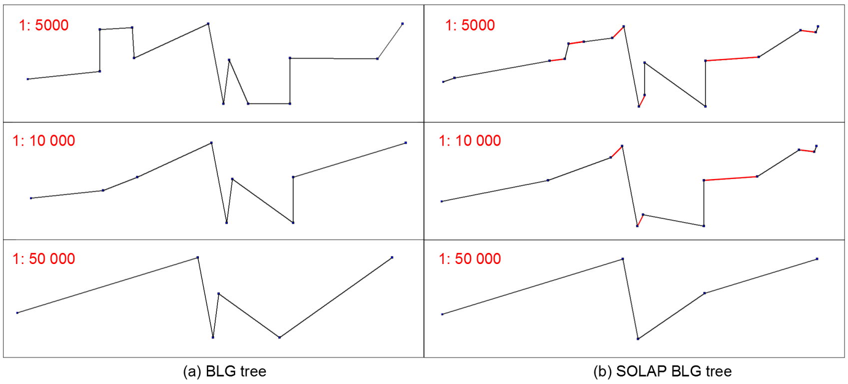

Fig. 10. Comparaison between the results of generalization by the SOLAP BLG tree and the BLG tree

5 Conclusions and future issues

This paper presented an on-the-fly generalization approach adapted to SOLAP applications. This process is intimately linked to the highly interactive applications in cartography such as web mapping, mobile mapping and SOLAP applications. But this latters, require the simultaneous consideration of cartographic and decisional aspects by integrating the measure in the process.

Our proposed approach is based on SOLAP BLG tree, it consists on adapting BLG tree structure, initially dedicated to cartographic generalization, to SOLAP. It focuses on linear objects (roads in our case study) and integrates the measure in order to adapt the level of detail that meets the decision-makers needs.

To validate our approach, we chose the road risk phenomenon as analysis subject, this is particularly due to its worldwide socio-economic impact. In addition, the use of map in the analysis of such phenomenon is of major interest for decision-makers because it is closely related to geographic information represented by the road object and the locality to which it belongs. In our experiments, we have highlighted the contribution of the proposed structure in the context of SOLAP through the various implemented functions such as importance function, and whose application has allowed preserving measures while providing cartographic perceptibility.

As future issues, we suggest : (1) improving the current solution by adapting the other generalization operators (as is smoothing, displacement, typification, exaggeration, etc.) to SOLAP applications. This will allow generating a better quality of maps and hence, improve the decision making process. (2) Adapt the generalization process to SDW by using another generalization approach, for example, the one, based on rapid generalization algorithms and (3) elaborate a comparative study between the implemented approaches according to some defined criteria in order to assess their effectiveness in a given context of use.

References

1. Chaudhuri, S., Dayal, U.: An overview of data warehousing and olap technology. ACM Sigmod record **26**(1) (1997) 65–74
2. Kimball, R., Ross, M., et al.: The data warehouse toolkit: the complete guide to dimensional modelling. Nachdr.]. New York [ua]: Wiley (2002)
3. Thomsen, E.: OLAP solutions: building multidimensional information systems. John Wiley & Sons (2002)
4. Rivest, S., Bédard, Y., Proulx, M.J., Nadeau, M., Hubert, F., Pastor, J.: Solap technology: Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data. ISPRS journal of photogrammetry and remote sensing **60**(1) (2005) 17–33
5. Malinowski, E., Zimányi, E.: Advanced data warehouse design: from conventional to spatial and temporal applications. Springer Science & Business Media (2008)
6. Bédard, Y.: Spatial olap. In: Forum annuel sur la RD, Géomatique VI: Un monde accessible. (1997) 13–14
7. Bimonte, S., Bertolotto, M., Gensel, J., Boussaid, O.: Spatial olap and map generalization: Model and algebra. International Journal of Data Warehousing and Mining (IJDWM) **8**(1) (2012) 24–51
8. Bereuter, P., Weibel, R.: Algorithms for on-the-fly generalization of point data using quadtrees. Proceedings AutoCarto 2012 (2012)
9. Weibel, R., Burghardt, D.: Generalization, on-the-fly. In: Encyclopedia of GIS. Springer (2008) 339–344
10. van Oosterom, P., Meijers, M.: Towards a true vario-scale structure supporting smooth-zoom. In: Proceedings of the 14th ICA/ISPRS workshop on generalisation and multiple representation, Paris. Volume 48. (2011)

11. van Oosterom, P., Meijers, M., Stoter, J., Šuba, R.: Data structures for continuous generalisation: tgap and ssc. In: Abstracting Geographic Information in a Data Rich World. Springer (2014) 83–117
12. Sarjakoski, L.: Conceptual models of generalisation and multiple representation. Generalisation of geographic information: Cartographic modelling and applications. Amsterdam, The Netherlands: Elsevier (2007)
13. Gaffuri, J.: Généralisation automatique pour la prise en compte de thèmes champ: le modèle GAEL. PhD thesis, Université Paris-Est (2008)
14. Stanislawski, L.V., Buttenfield, B.P., Bereuter, P., Savino, S., Brewer, C.A.: Generalisation operators. In: Abstracting Geographic Information in a Data Rich World. Springer (2014) 157–195
15. Lehto, L., Sarjakoski, L.T.: Real-time generalization of xml-encoded spatial data for the web and mobile devices. International Journal of Geographical Information Science **19**(8-9) (2005) 957–973
16. Foerster, T., Lehto, L., Sarjakoski, T., Sarjakoski, L.T., Stoter, J.: Map generalization and schema transformation of geospatial data combined in a web service context. Computers, Environment and Urban Systems **34**(1) (2010) 79–88
17. GiMoDig: Geospatial info-mobility service by real-time data-integration and generalisation (2001)
18. Lejdel, B., Kazar, O.: Genetic agent approach for improving on-the-fly web map generalization. CoRR (2012)
19. Van Oosterom, P.: The reactive-tree: A storage structure for a seamless, scaleless geographic database. In: AUTOCARTO-CONFERENCE-. Volume 6., ASPRS AMERICAN SOCIETY FOR PHOTOGRAMMETRY AND (1991) 393–393
20. Van Oosterom, P.: Variable-scale topological data structures suitable for progressive data transfer: The gap-face tree and gap-edge forest. Cartography and Geographic Information Science **32**(4) (2005) 331–346
21. Douglas, D.H., Peucker, T.K.: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica: The International Journal for Geographic Information and Geovisualization **10**(2) (1973) 112–122
22. Bereuter, P., Weibel, R.: Real-time generalization of point data in mobile and web mapping using quadtrees. Cartography and Geographic Information Science **40**(4) (2013) 271–281
23. Bencherif, H., Boubakour, F., Belkacem, N.: Les accidents de la route dans les médias de masse en algérie. du traitement de l'information à sa diffusion. Communication. Information médias théories pratiques **30**(1) (2012)
24. Amieur, K.D., Frihi, I., Boukhalfa, K., Alimazighi, Z.: De la conception d'un entrepôt de données spatiales à un outil géo-décisionnel pour une meilleure analyse du risque routier. (2013)
25. Derbal, K., Ibtissem, F., Boukhalfa, K., Alimazighi, Z.: Spatial data warehouse and geospatial decision making tool for efficient road risk analysis. In: Information and Communication Technologies for Disaster Management (ICT-DM), 2014 1st International Conference on, IEEE (2014) 1–7