Toward the generalisation of cartographic mashups:
Taking into account the dependency between the thematic data and the reference data throughout the process of automatic generalisation

Kusay Jaara, Cécile Duchêne, Anne Ruas
COGIT labaratory, IGN France
2 avenue Pasteur - 94160 Saint-Mandé, FRANCE
kusay.jaara@ign.fr, cecile.duchene@ign.fr, anne.Ruas@ign.fr

1. Introduction

In Computer science, a mashup is a web page or an application that combines data or functionality from two or more external sources to create a new service (Wikipedia). “Mapping mashup" means to combining data (pictures, videos, news…etc) with a mapping application’s locating service like Google maps.

In our article, we use the term “cartographic mashups” for referring to group of geographic data layers which may come from more than one source, these layers are superimposed and are related semantically. In addition, these layers could be thematic or topographic.

The integration problems of superimposition have been discussed in precedent works, but the complexity has increased with the distribution of data throughout the web and the Volunteered Geographic Information (VGI).

With the use of geographic data in so many applications, a lot of producers make the data available through the web, in order to make it accessible to a maximum number of people. These data have different quality, scale, degree of details and specifications. The data may be thematic or topographic, the thematic data are related to one theme and has a localisation(point or area), topographic data serve at locating the thematic data.

One example of cartographic mashups is geoportals where data collected from several producers can be co-visualised, another example is the web applications that enables users to create their own customized maps.

This paper intruduces a beginning research study about the generalization of mashup maps which is composed of several layers, among which some “reference” topographic layers and one or several “thematic” layers.

We need to perform generalization on these mashup data:

- In order to increase the consistency by changing the scale or the level of detail of layers.
- In order to decrease the level of detail of one or more layers, to adapt it to the context of use, the intended display scale and the associated graphical constraints.

In our work, we focus on the co-visualisation of geographic data in the case of combing topographic and thematic layers of different sources and possibly different level of detail, quality and where we may need to perform generalisation to reduce the scale of visualisation.

In our article, we work on geographic data that could be obtained from the web or offline from a data producer.

After the problem statement §2, we start to illustrate the problem by one example §3, next we present our analysis and first thoughts on referencing thematic data with respect to topographic data §4, finally we present related works §6.

2. Problem statement

We will use the terms “themes” and “layers”. Theme means the category of cartographic data, for example the theme “transport” which includes the road network and rail roads layers. One theme may contain more than one layer. Every layer comes from one source and has one theme.

The problems we encounter in our case are listed in two categories:

Problems of integration

- The problem of shifts due to the variation of resources, or the different representation of reality.
- The semantic relations that are not always respected (road accidents has to be on the roads)
- Data redundancy where we can have more than one representation of the same object.

Problems related to generalization:

- The difference of level of detail, which can be different from one layer to another, so we may need to aggregate or reduce the level of detail to ensure the consistency between thematic and topographical data.
- The level of detail may vary from one zone to another in the same theme. In some cases we cannot integrate the date when the difference of level of detail is important, for this reason the generalization process could be useful during the process of integration.
- The need to do the generalization in order to visualise at a smaller scale.
- The level of detail of visualisation has to be consistent and correspondent to the type of the map, this consistency is an important condition e.g. for the INSPIRE Directive (Toth, 2007)

In fact, the way we manipulate the topographic data has to take into consideration the relations between thematic and topographical data, for example, a bus line pass by a narrow road which has to be preserved throughout the generalization because a bus line passes by it. The issue in the level of detail is presented in the thematic data as well as the topographic data. We may also reduce the details in both of them at the same time.

3. Illustration example:

In order to illustrate the issue that we want to study, we have done a first analysis on a simple case study

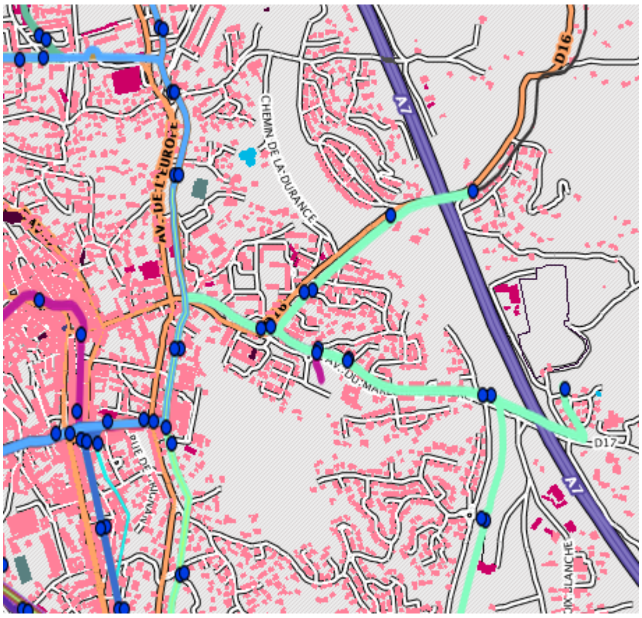


Figure 1: The bus lines in the region of Salon-de-Provence, France
source : Geoportal of France

The above figure shows a map visualised on a the French geoportal (www.geoportal.fr), it consists an extract of BDTopo(produced by IGN France) region of Salon-de-Provence, on which we superimpose the bus lines coming from another source

In this example, we would like to be able to do the following processes automatically:

- Make the reasoning for that "the bus lines follow the road network".
- Match the thematic data (the bus lines) with the topographic data of road network (here the bus lines are shifted).
- Create a representation more consistent from the two components(the bus lines and the road network)
- Throughout the generalization process when the map is zoomed out, we have to ensure that the characteristic points of bus lines (the starting point and the end of bus line, the stations) stay in the same place compared to characteristic points of road network, and to other objects (for example, administrative buildings). We want to conserve the relations between thematic data and topographic data, sometimes we may want to exaggerate them.

4. Referencing thematic data on topographic data: first thoughts

In order to conserve (or exaggerate) the relations between the thematic data and topographic data, we need a clear method to link them. The “reference system” attaches the thematic data to the topographic data.

This system contains two components:

- Semantic component: e.g. the bus station is related to the road.
- Spatial component : includes the relative position topologic(e.g. the bus station is on the road) and also in term of the metric position which means the fine referencing objects(e.g. Distance along the road)

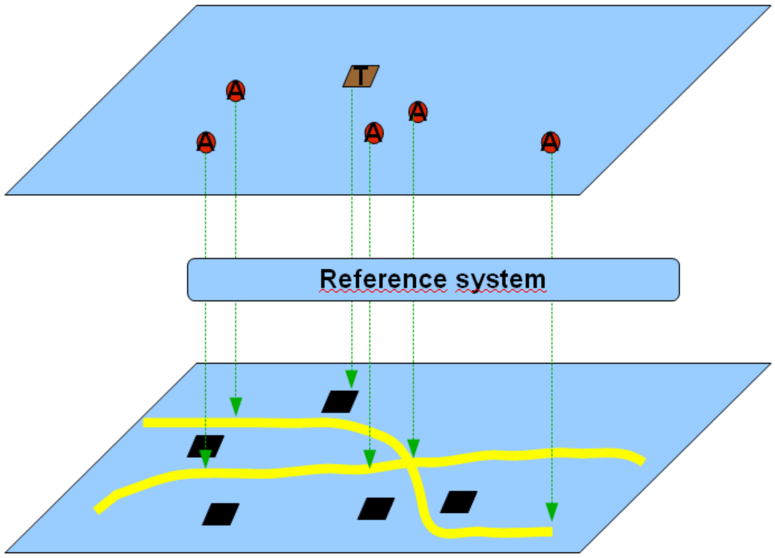


Figure 2: The referencing of thematic data, T=touristic building A=accident

5. First case study:

A contact was made with the The Technical Study Center of Equipment (CETE) of Aix department. This example shows a practical use of the topic. In addition, the example gave us a first idea for a model that serves for hooking the thematic data on the reference topographic data. The CETE in France working within the Ministry of Equipment conducts research and develops applications in many areas like transport, infrastructure and bridges.Part of their work is to provide tools that help in managing the road network (security rating, accidents, signs, traffic lights…). In addition, they provide tools which helps in analyzing this information.

The professional data (which is considered a thematic data for us) is the situation of every 200m of the road, the accidents, traffic lights, signs…etc.For the localization of their data, their system was relying on Physical objects (same as milestones) located beside the roads every one kilometer (with possible shift), the location is detected using one dimension, which was the distance from the milestones.

They added the geometry to their GIS system by linking their database to existing geographic databases. They decided in 2009 to make their database independent from the position system, so they redefined the structure to store the geometry of every object with the geometry source. The advantage of such a structure is the possibility to have more than one source of geometry in their database which serves as reference (e.g. road topographic data)

They also extended the concept of physical object by introducing the concept of the localisation points, which could be the mile stone, a house address of any recognized point of the road.

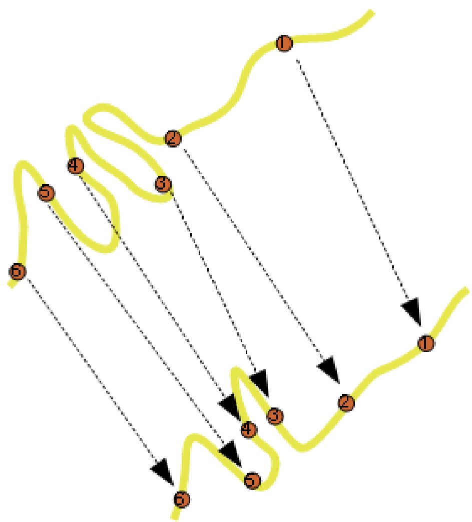
From an analytical point of view, they have a successful model for hooking punctual or linear thematic data on the topographical data which is considered as support layers.

In their model, the thematic data has a strong semantic relation to reference data. The reference data can be changed, but then we have to relocate the thematic data on the new reference data.

They use certain points, identify them on the reference layer, and then use them to mark the position of thematic data. By using this structure, it makes it easier to change the support data, because only the localisation points have to be located on the new support data.

In fact, their model is conceptually correct, but still has a disadvantage. The localization is not well expressed in the thematic layer. That is considered a real problem, because they use localisation points to locate the thematic data on the topographic layer, we cannot always locate the localisation points precisely. We just have its distance from the starting point of road, which is not enough to precisely locate the thematic data on any geographic database.

The problem they want to solve is that they may need to change the reference data, but the system to relocate the thematic data is not sufficiently precise. The relocating is done using the distance measured on the road between localisation points and the distance between them according to the reference data. The relocating is done using the rational way which is not always right.



up : thematic data related to a reference database
bottom : thematic data related to another reference database
Figure 3: thematic data transferred from one database to another

In order to locate the thematic data more precisely, we propose completing the model by inserting characteristic points to the system, these points include the inflection points, turning points or other points that can be recognized even with the change of the geographical database or even the scale.

The information we want to store of these points is not just the distance from the road starting. In fact we extract the spatial relations of our points with the road network and even with other objects if possible.

6. Related work

The subject never been discussed in the same way, some previous work aims to solve special cases of our subject, others could be a part of the whole solution. Regarding the problem of integration between layers, we will need the integration techniques, from which we have the spatial integration and matching (Olteanu 2007) (Sheeren 2009 et al.)

We can get better results of integration by extracting semantic relations, for this we can extract ontologies from specifications then align them (Abadie 2009a).

We can use the specifications to identify the heterogeneities between databases (Abadie 2009b) and that will also help in reducing the redundancy.

We can also use the techniques used for manipulating the multiple representation databases like we have in (Vangenot and al. 2002) where unique name and container is used for the multiple representations and identify the specifics of each representation using the stamping techniques. His model could be useful to manage the mashups that has layers of different level of detail.

As we have relational conditions that has to be satisfied between objects of different layers, we can represent them by creating relational constraints like in (Duchene 2007)

In order to control the generalisation by the thematic data, we can use the model of (Gaffuri 2008 and al.) which is a multi-agent system where the objects have relations with fields. In our case, we would use the thematic data as fields. In his model, the objects and the fields can be modified in order to answer the constraints.

We can use Constraint tGAP method (Dilo and al., 2009) to get a better generalization result in the case of having two topographic layers with two different scales, and when the requested scale is between the scales of the two layers.

7. Conclusion and Perspective

In this article, we presented the beginning of a study concerning how to take into account the thematic data throughout the generalization process in the case of cartographic mashups. The study is mainly regarding the case of having the two types of data “topographic” and “thematic” among the mashup layers.
The processing needed for the co-visualisation of mashups, require the integration process. We may need the generation process, because when the visualisation scale is reduced, a particular attention has to be done to maintain or even to enhance the relationships between thematic and topographic components.
Our further work includes the proposition of a generic model for such a referencing system. Then we will study how to respect these relations throughout the generalisation process.

References

(http://en.wikipedia.org/wiki/Mashup_(web_application_hybrid)

Abadie N., "*Schema Matching Based on Attribute Values and Background Ontology*", 12th International Conference on Geographic Information Science (AGILE'09), Hanovre (Germany), June 2009

Abadie N.*"Formal specifications to automatically identify heterogeneities"* AGILE Pre-Conference Workshop (Challenges in Geospatial Data Harmonisation) Hannover (Germany), June 2009

Olteanu A-M., 2007, "*A multi-criteria fusion approach for geographical data matching*", 5th International Symposium on Spatial Data Quality (ISSDQ'07), June 2007, Enschede (Netherlands)

Buttenfield B. Wolf E. *"The Road and the River Should Cross at the Bridge" Problem:Establishing Internal and Relative Topology in an MRDB"* 10th ICA Workshop on Generalization and Multiple Representation, August 2007, Moscow

Buttenfield B. Stanislawski L. Brewer C. "*Multiscale Representations of Water: Tailoring Generalization Sequences to Specific Physiographic Regimes*" GIScience 2010 Short Paper Proceedings, Zurich, Switzerland.

Dilo A. van Oosterom P. Arjen H. "*Constrained tGAP for generalisation between scales:the case of Dutch topographic data*" Computers, environment and urban systems vol. 33, no5, pp. 388-402

Duchêne C.*"The CartACom model : a generalisation model for taking relational constraints into account",* 7th ICA Workshop on Generalisation and Multiple Representation, 20-21 August 2004, Leicester (UK)

Egenhofer M. Frannk A. Jachson j. "*a topographical data model for spatial databases"* Symposium on the Design and Implementation of Large Spatial Databases,1989, Santa Barbara, CA

Gaffuri J. Duchêne C. Ruas A. "*Object-field relationships modelling in an agent-based generalisation mode*" 11th ICA workshop on Generalisation and Multiple Representation, 20-21 June 2008, Montpellier, France

Mustière S. Abadie N. Laurens F. 2007 "*Appariement de schémas de BD géographiques à l'aide d'ontologies déduites des specifications*", Atelier DECOR (passage à l'échelle des techniques de découverte de correspondances), 22-23 janvier, Namur (Belgique), pp 22-27

Sheeren D., Mustière S., Zucker, J.-D., 2009, "*A data-mining approach for assessing consistency between multiple representations in spatial databases*", International Journal of Geographical Information Science, Vol 23, Issue 8, August 2009, pp 961–992

Stigmar H. "*Merging route data and cartographic data*", ICA workshop on generalisation and multiple representation, 2004, Leicester, England

Toth K. "*Data consistency and multiple-representation in the European Spatial Data Infrastructure",* 10th ICA Workshop on Generalisation and Multiple Representation, *2007, Moscow*

Vangenot C. Parent C. Spaccapietra S. "*Modelling and Manipulating Multiple Representations of Spatial Data"* Symposium on Geospatial Theory, Porcessing and Application, SDH'02. Proceedings of the 10th International Symposium on Spatial Data Handling, Ottawa, Canada, July 9-12, 2002.