

Chapitre X

La généralisation à la volée

1.1. Introduction

Pendant plusieurs années, la cartographie fut principalement consacrée à des produits statiques (cartes papiers et cartes numériques statiques). Avec l'arrivée du Web, plus récemment des dispositifs mobiles (ex. assistants personnels ou *PDA* : *Personal Digital Assistant*) et de la technologie SOLAP (*Spatial On-Line Analytical Processing*), la cartographie a pris un nouveau tournant et doit maintenant satisfaire de nouvelles exigences. De par ces nouveaux médias, l'information géographique est dorénavant accessible à une plus grande quantité d'individus, engendrant ainsi une plus grande variété de besoins à satisfaire. En effet, chaque utilisateur possède ses propres préférences en ce qui concerne les données devant être visualisées. De plus, ces médias offrent une interactivité n'ayant pu être offerte jusqu'ici via de simples cartes statiques. L'utilisateur a maintenant l'opportunité de naviguer parmi les données spatiales en modifiant par exemple les thèmes ou le niveau de détail visualisés. Cette nature dynamique et interactive requiert toutefois des temps de réponse quasi-instantanés. Par conséquent, les changements d'échelle doivent être effectués en temps réel, exigeant ainsi une généralisation à la volée (*on-the-fly generalization*). Cette généralisation doit permettre de générer des produits cartographiques à des échelles arbitraires (définies par l'utilisateur) et ce, sans aucun délai notable. La plupart des approches proposées jusqu'à maintenant afin de répondre à ce besoin sont basées soit sur le développement et l'utilisation d'algorithmes de généralisation automatique (approche orientée-traitements), soit sur une base de données supportant différents niveaux de détail (approche orientée-

2 La généralisation à la volée pour le web

représentations). Peu d'entre elles proposent de combiner les deux approches, ce qui pourrait, selon nous, mener au développement de systèmes plus efficaces pour rencontrer ces nouveaux besoins.

La suite du présent chapitre vise à cerner cette possibilité. Nous définissons dans un premier temps ce qu'est la généralisation à la volée et discutons des deux approches actuelles pour son implantation. Nous présentons par la suite deux applications pour lesquelles ce processus s'avère essentiel soit, la cartographie interactive sur le Web et l'exploration interactive des données spatiales (SOLAP). Pour chacune des applications, les contraintes apportées à la généralisation à la volée sont présentées. Puis, nous faisons une révision plus détaillée de l'état de l'art des algorithmes de généralisation utilisés dans l'approche orientée-traitements ainsi que des solutions actuelles de l'approche orientée-représentations. Nous terminons finalement avec la présentation de deux solutions orientées vers la combinaison des deux approches.

1.2. Définition et applications de la généralisation à la volée

1.2.1. Définition de la généralisation à la volée

La généralisation à la volée peut être vue comme étant un sous-ensemble d'une tendance plus générale nommée **cartographie sur demande** (*on-demand mapping*). Celle-ci permet de générer des produits cartographiques en fonction des besoins des utilisateurs. Que les cartes soient produites en ligne ou non, en temps réel ou en quelques jours et que les produits soient sur support papier ou numérique n'a que peu d'importance. La principale caractéristique réside dans le fait que l'utilisateur spécifie lui-même *quand* et *comment* la carte doit être produite. Ainsi, une carte papier créée à la demande et selon les spécifications d'un utilisateur (ex. *Ordnance Survey Landplan* [GOW 97]) peut être qualifiée de « carte sur demande » tout comme une carte indiquant la position d'un utilisateur à l'écran d'un assistant personnel.

Contrairement à [JON 00], pour qui la généralisation à la volée est synonyme de cartographie sur demande, nous croyons que ces deux termes signifient des concepts suffisamment différents pour ne pas être confondus. Selon nous, les exigences de la cartographie et de la généralisation à la volée¹ sont relativement plus difficiles à

¹ « Effectuée en temps réel » est un synonyme de « à la volée ». [OOS 95] définit la généralisation cartographique à la volée comme un processus automatique permettant la création instantanée et temporaire d'un ensemble de données généralisé, servant exclusivement pour la visualisation et qui n'est pas emmagasiné ou utilisé à d'autres fins.

satisfaisant. Premièrement, les cartes doivent être générées en temps réel, ce qui rend le processus beaucoup plus contraint par le facteur temps. Deuxièmement, étant donné que les traitements en temps réel nécessitent des médias graphiques dynamiques, les produits résultants seront indubitablement des cartes numériques sujettes à des contraintes techniques plus strictes.

Ainsi, nous pouvons définir la généralisation à la volée comme étant : la création, en temps réel, d'un produit cartographique approprié à son échelle et à son but, suite à la demande d'un utilisateur et effectuée à partir d'une base de données à plus grande échelle. Les principales caractéristiques sous-jacentes à ce processus sont :

- un ensemble de données, à plus petite échelle (généralisé), est généré temporairement à des fins de visualisation à partir d'une base de données possédant un niveau de détail plus fin;
- la visualisation des données doit satisfaire les préférences de l'utilisateur ainsi que les spécifications techniques d'affichage;
- l'échelle est arbitraire et non prédéfinie;
- le processus de généralisation est effectué automatiquement et il est impossible de vérifier le résultat avant l'affichage;
- la carte résultante doit s'afficher en quelques secondes puisque l'utilisateur ne désire pas attendre [FER 01];
- pour le Web et les dispositifs mobiles, la largeur de la bande passante (*network bandwidth*) pose un problème supplémentaire quant au volume de données transmis ainsi qu'au type d'architecture logicielle possible (client lourd, client léger, applet, page statique).

Les points mentionnés ci-haut démontrent les défis de la généralisation à la volée, particulièrement pour la cartographie sur le Web. Cependant, il doit être mentionné que la généralisation à la volée n'est pas seulement restreinte aux applications Web mais peut également s'avérer essentielle pour les applications de cartographie interactive ou mobile ainsi que pour les outils appuyant la découverte de connaissance géographique (*Geographic Knowledge Discovery*) tels que les outils SOLAP.

En principe, on distingue deux approches fondamentales pour l'implantation de la généralisation en temps réel. La première approche, orientée-représentations, repose sur la saisie ou le précalcul (voire même la production manuelle) de représentations multiples dans une base de données multiéchelle supportant

4 La généralisation à la volée pour le web

différents niveaux de détail. Bien que cette solution soit techniquement plus simple, moins exigeante au niveau des traitements et réalisable avec les technologies actuelles, elle est également la moins flexible car restreinte à l'exploitation des représentations offertes par le système. Au contraire, la deuxième approche, que l'on dit orientée-traitements, est basée sur des processus de généralisation effectués en temps réel à partir d'un jeu de données uniéchelle à représentation simple des objets. Quoiqu'elle requière plus de traitements et qu'elle soit d'une plus grande complexité technique, cette seconde approche démontre davantage de flexibilité car elle n'est pas restreinte à des niveaux de détail prédéfinis. Par contre, l'état actuel des connaissances pour cette approche ne permet de mettre en place que des systèmes dont les limites apparaissent rapidement (ex. objets complexes, combinaison de thèmes différents). Il nous semble cependant qu'il soit possible de combiner les forces de chaque approche, soit la maturité technologique et les temps de traitement d'une part (ce qui semble favorable à l'approche orientée-représentations) et la flexibilité de génération de produits ainsi que la qualité cartographique d'autre part (qui sont les forces de l'approche orientée-traitements).

De manière réaliste, les applications actuelles de la généralisation en temps réel reposent sur l'approche orientée-représentations, i.e. sur des bases de données multiéchelles (avec différents niveaux de détail) car aujourd'hui, la généralisation à la volée se situe surtout au stade de recherche. Par contre, une meilleure combinaison des deux approches permettrait d'obtenir des résultats optimaux avec des systèmes plus efficaces.

1.2.2. Définition et contraintes de deux applications qui requièrent une généralisation à la volée

1.2.2.1. La cartographie sur le Web

Parallèlement à la croissance rapide de l'Internet, la cartographie sur le Web a fait l'objet d'une évolution fulgurante. D'un point de vue quantitatif, le Web est devenu le plus important média pour la diffusion de cartes géographiques [KRA 01]. Selon [PET 99], environ 40 millions de cartes sur le Web sont utilisées par jour et ce, à travers le monde entier. La cartographie Web offre deux principaux avantages : accessibilité et actualité [ELZ 01]. L'accessibilité provient du fait que l'utilisateur a accès à une grande quantité d'information et de cartes en tout temps et en tout lieu², ou presque. L'actualité réfère au fait que les mises à jour au niveau de la base de données peuvent être immédiatement reflétées sur les cartes Web. Le manque d'actualité était d'ailleurs l'une des principales lacunes de la cartographie

² En cartographie mobile (i.e. la cartographie via des dispositifs mobiles), les services sont en effet disponibles en tout lieu.

traditionnelle; au moment de leur publication, les cartes n'étaient souvent plus à jour étant donné la durée du processus conventionnel de production cartographique.

Au commencement de son évolution, la cartographie Web était restreinte à de simples cartes statiques de format matriciel (GIF ou JPG). Depuis, les capacités techniques (représentations graphiques, largeur de la bande passante) d'une part et les exigences des utilisateurs d'autre part n'ont cessé de croître, provoquant ainsi une évolution constante vers la cartographie Web sur demande et à la volée. L'état de l'art actuel des sites Web commerciaux offrant des fonctionnalités de cartographie Web est parfaitement illustré par des sites tels que MapQuest³ et MS Expedia Maps⁴. Pour des cartes à des fins plus spécialisées telles que les cartes de repérage ou de recherche de parcours, l'utilisateur a la possibilité de choisir la fenêtre qui est affichée (i.e. la localisation), l'échelle d'affichage et possiblement les thèmes de données devant être affichés. Les cartes sont ensuite extraites à partir du niveau d'échelle approprié de la base de données multiéchelle (composée de plusieurs niveaux d'échelle préétablis qui ne sont pas interconnectés). La carte prédéfinie est par la suite envoyée sous format matriciel ou vectoriel (via des technologies telles que GeoMedia WebMap, Jmap) au client Web. La flexibilité de la cartographie sur demande est relativement faible concernant le choix des thèmes ou de l'échelle de la carte. Toutefois, si l'utilisateur demeure dans le contexte prédéfini du serveur cartographique (ex. tracés d'itinéraires), les résultats sont acceptables comme en témoigne le nombre d'accès quotidien à ce genre de site.

Indépendamment de l'approche utilisée pour l'implantation, plusieurs contraintes techniques influencent la représentation des cartes sur le Web. Parmi les plus importantes figurent (pour plus de détails voir [ARL 99]):

- **le taux de transfert** : la cartographie Web doit s'accommoder des taux de transfert plutôt limités. Actuellement, les taux de transfert varient entre 52 et 128 kbps pour les modems téléphoniques, entre 1 et 1.5 mbps pour les modems câbles et se situe entre 30 à 144 kbps pour les technologies sans fil;
- **la qualité et la profondeur de la couleur** : des profondeurs allant de 16 à 24 bits sont utilisées couramment aujourd'hui et sont suffisantes pour la plupart des applications. L'affichage des couleurs peut toutefois différer d'un écran à l'autre (i.e. la même couleur semblera différente sur des écrans différents). Cependant, en ce qui concerne les assistants personnels et les téléphones cellulaires, encore plusieurs sont monochromes;

³ www.mapquest.com

⁴ www.expedia.com

6 La généralisation à la volée pour le web

- **la résolution de l'écran** : bien que la résolution des écrans de micro-ordinateurs se soit améliorée continuellement de sorte qu'une résolution de 1024 par 768 pixels soit couramment utilisée, les écrans d'assistants personnels n'offrent qu'une résolution d'un quart de VGA, i.e. de 320 par 240 pixels. Il est clair que pour les années à venir, nous devons composer avec des résolutions infiniment moins bonnes que celles des produits papiers.

Une difficulté supplémentaire pour la cartographie Web provient du fait que ces contraintes techniques sont incontrôlables par le concepteur de la carte, les caractéristiques précises de l'affichage et du transfert de données pouvant varier significativement d'un utilisateur à un autre. Également, en raison de la portée incommensurable du Web et du fait que quiconque puisse y accéder (à moins que des restrictions soient imposées délibérément), il est extrêmement difficile d'émettre des suppositions quant aux futurs utilisateurs de la carte relativement à leur capacité de lecture cartographique, à leur niveau d'expertise, etc.

Considérant que toutes les applications Web reposent sur une architecture client/serveur, la répartition de la charge entre le serveur et le client devient donc une question majeure. Dans la plupart des cas actuels, la carte demandée est générée par le serveur et transférée au client en tant qu'image matricielle. Cette approche côté-serveur possède l'avantage de permettre au client Web d'être « léger », en nécessitant que peu ou pas de fonctionnalités spécialisées (ex. un navigateur Web dépourvu de plugiciels). Le micro-ordinateur n'a donc pas à être puissant étant donné que presque tous les calculs s'effectuent au niveau du serveur. De plus, ce type de solution permet d'avoir un contrôle absolu sur les caractéristiques et la sécurité (accès aux données) des cartes Web générées (en fait, l'utilisateur ne peut modifier ni les caractéristiques de l'affichage ni les données de manière locale). Cette approche repose essentiellement sur le transfert des données entre le serveur et le client, puisque chaque changement effectué par l'utilisateur (ex. zoom et pan) engendre une nouvelle requête au serveur et un nouveau transfert de la carte vers le client.

L'approche alternative consiste à introduire plus de fonctionnalités du côté client, ce dernier devenant généralement un client « lourd » (typiquement un applet Java ou un plugiciel spécialisé d'un navigateur Web) pouvant traiter certaines requêtes localement (zoom, pan, changement de thèmes, modification de la symbologie, etc.). Cette approche permet de diminuer le nombre de requêtes au serveur, d'augmenter la flexibilité de l'affichage local en plus d'accroître l'interactivité. Toutefois, cette stratégie implique que non seulement la carte demandée mais également toutes autres informations (ex. niveaux de détail additionnels à des fins d'agrandissement ou de réduction de l'échelle d'affichage)

doivent être téléchargées au client. Généralement, des solutions de ce genre ne sont utilisées qu'avec des formats vectoriels tels que SVG (*Scalable Vector Graphics*) et Flash ou encore, avec des formats vectoriels propriétaires de certaines technologies (ex. *GeoMedia WebMap*, *JMap*).

La récente émergence des formats vectoriels polyvalents (en particulier SVG) vont, espérons-le, rendre possible la combinaison des avantages des deux approches, en effectuant par exemple le prétraitement de la généralisation à la volée au niveau du serveur et en offrant certaines fonctionnalités du côté client (zooms et autres contrôles). Une répartition des opérations de généralisation peut également se faire entre les postes clients et le serveur lorsqu'une structure vectorielle est utilisée.

Somme toute, il y a deux avantages à l'introduction de la généralisation à la volée au sein de la cartographie Web :

- elle rend possible une cartographie mieux adaptée aux besoins spécifiques des utilisateurs. Ces derniers peuvent définir leurs préférences notamment en ce qui concerne les thèmes à afficher, l'échelle ainsi que la symbologie à utiliser tout en se prévalant d'une carte de qualité cartographique adéquate;
- la généralisation à la volée accélère la création de la carte demandée. Pour une solution côté-serveur, des algorithmes plus puissants peuvent être utilisés alors que pour une solution côté-client, la vitesse et la versatilité de l'affichage sont plus grandes.

Par conséquent, la cartographie Web apporte une toute nouvelle signification à la généralisation. Considérant le développement technologique rapide des applications mobiles telles que les services basés sur la localisation (*location-based services*), cette signification prendra encore plus d'importance du fait que les contraintes techniques des applications sans-fil sont encore plus contraignantes (taux de transfert, résolution et taille de l'écran, couleur, etc.).

1.2.2.2. Les outils SOLAP

La récente démocratisation de l'information géographique a engendré une augmentation considérable du volume de données emmagasiné au sein des bases de données spatiales. Afin d'exploiter efficacement toute la richesse de cette information, il devient nécessaire de développer de nouveaux outils dédiés à l'analyse et à l'exploration de données spatiales.

Les outils SOLAP sont spécifiquement conçus afin de supporter les processus d'analyse spatio-temporelle. De par leur grande facilité d'interrogation, ils appuient la prise de décision stratégique. En réalité, l'utilisateur d'un tel outil n'a pas à

8 La généralisation à la volée pour le web

maîtriser de langages d'interrogation (ex. SQL) de même qu'il n'a pas à connaître a priori la structure sous-jacente de la base de données (nom des tables, relations, etc.). Plus formellement, un outil SOLAP peut être défini comme étant « une plate-forme visuelle supportant l'exploration et l'analyse spatio-temporelle simples et rapides des données selon une approche multidimensionnelle à plusieurs niveaux d'agrégation via un affichage cartographique, tabulaire ou en diagramme statistique » [BED 97].

Leur structure multidimensionnelle est optimisée pour les processus d'analyse puisqu'elle reflète la façon dont les gens perçoivent naturellement leurs données. Cette structure est basée sur les concepts de dimensions, de mesures et de faits [HAN 98 ; THO 99 ; RIV 01]. Une **dimension** représente un thème d'analyse (ex. temps, région, produit) et contient des membres qui sont organisés hiérarchiquement selon le **niveau d'agrégation**. Par exemple, une dimension temps peut contenir trois niveaux hiérarchiques tels que *jour*, *mois* et *année* alors qu'une dimension région géographique peut être composée des niveaux *ville*, *province* et *pays*. Une **mesure** est une donnée généralement numérique sur laquelle portent les analyses (ex. total des ventes). Le **fait** représente la valeur d'une mesure résultant d'un croisement de plusieurs membres appartenant chacun à une dimension particulière (ex. le total des ventes du produit X, pour la région Y et pour l'année Z). Finalement, la construction multidimensionnelle résultant de la conjonction de plusieurs dimensions forme ce qu'il convient d'appeler un « **hypercube** » dans lequel chaque cellule est définie par un seul membre de chaque dimension et représente la valeur d'une mesure à l'intersection de ces dimensions.

La navigation au sein d'un outil SOLAP s'effectue intuitivement par le biais de différents opérateurs tels que le pivot (*swap*) qui permet d'interchanger deux dimensions, le forage (*drill-down*) et le remontage (*drill-up*) qui permettent d'analyser les données avec plus ou moins de détails respectivement et le forage latéral (*drill-across*) qui permet d'analyser un autre thème pour un même niveau de détail. Ainsi, via ces opérateurs, l'utilisateur peut facilement naviguer parmi un jeu de données en interchangeant les thèmes d'analyse et en augmentant ou diminuant le niveau de détail avec lequel les données sont affichées.

Une des caractéristiques fondamentales des outils SOLAP est qu'ils permettent une analyse **rapide** des données afin qu'un utilisateur puisse conserver son schème de pensée lors de la navigation dans le système. Par conséquent, à l'instar de la cartographie Web, la principale contrainte imposée par ce type d'outils est que les temps de réponse doivent être quasi-instantanés. À cet égard, l'aspect spatial des données constitue donc un défi majeur pour une implantation efficace [STE 00]. En effet, alors qu'il est relativement facile et rapide d'effectuer une opération descriptive (ex. sommation des ventes d'un produit pour une année entière, moyenne des températures pour une région donnée), il en est autrement avec les opérations

spatiales (ex. agrégation des régions selon la quantité de neige reçue, agrégation des départements en régions administratives). Celles-ci requièrent bien plus que de simples opérations mathématiques et engendrent généralement des temps de traitement significatifs. Il serait donc avantageux de stocker explicitement le résultat de certaines opérations spatiales (les plus fréquemment utilisées) afin d'accélérer les temps de réponse. Toutefois, tous les résultats ne pouvant être stockés a priori (ce qui engendrerait un volume de données trop élevé), certaines opérations doivent être effectuées à la volée [HAN 98] via des processus de généralisation automatique. Ceux-ci devront engendrer des temps de réponse rapides afin de ne pas interrompre l'utilisateur lors de la navigation dans le système.

De plus, afin de faciliter le processus d'analyse, il est essentiel de conserver une certaine cohérence sémiologique entre les représentations à différents niveaux de détail d'un même phénomène. Par exemple, les représentations géométriques détaillée et simplifiée d'un même bâtiment doivent partager les mêmes variables visuelles tout comme la zone urbaine, créée suite à l'agrégation de certains bâtiments, doit respecter la sémiologie graphique de ceux-ci. Cette continuité sémantique est nécessaire afin d'identifier facilement les objets d'une même famille. Les mécanismes de généralisation automatique doivent donc prendre en considération la sémiologie graphique des éléments à généraliser.

Par ailleurs, considérant qu'un outil SOLAP se compose de deux volets, l'un sémantique et l'autre cartographique, la granularité sémantique d'un objet devrait être adaptée à sa granularité géométrique pour une cohérence accrue. Ainsi, la généralisation cartographique d'un objet devrait de surcroît engendrer une généralisation sémantique. Par exemple, un objet défini comme étant une maison à grande échelle (avec une géométrie complexe) pourra être défini comme étant une construction à plus petite échelle (avec une géométrie simplifiée). Dans un tel contexte, la généralisation ne doit donc pas s'effectuer uniquement au niveau de la géométrie des objets mais également au niveau de leur sémantique.

Conséquemment, ces contraintes doivent être prises en considération lors de l'implantation des processus de généralisation à la volée dans un contexte SOLAP. Ceux-ci devront s'effectuer rapidement et tenir compte des composantes géométriques aussi bien que sémantiques et sémiologiques des données⁵.

Notons finalement que les outils SOLAP sont dédiés beaucoup plus à des analyses de type exploratoire que des analyses métriques ou topologiques telles que retrouvées dans un SIG conventionnel. En fait, seul le niveau le plus détaillé sert à des analyses plus précises. Les autres niveaux, présentant des données agrégées ou

⁵ Voir le chapitre intitulé « La métastructure VUEL et la gestion des représentations multiples » de ce même volume.

résumées servent plutôt à déceler des tendances générales. Ainsi, à ces niveaux, la clarté de visualisation est généralement plus importante que l'exactitude métrique.

1.3. État des connaissances

1.3.1. État de l'art de l'approche orientée-traitements

Tel que mentionné précédemment, la généralisation à la volée est une tâche pour laquelle le temps joue un rôle critique. Par conséquent, les algorithmes de généralisation appropriés à ce processus doivent être rapides et/ou doivent être supportés par des attributs ou des structures de données précalculés. En principe, tous les algorithmes de généralisation actuels qui s'exécutent selon un temps linéaire ou logarithmique peuvent être utilisés à des fins de généralisation à la volée. Parmi ce type d'algorithme figurent les algorithmes de sélection qui reposent exclusivement sur des attributs précalculés tel que le *Horton stream ordering scheme*, utilisé pour la sélection de réseaux fluviaux [RUS 90]. Un autre exemple d'algorithme de sélection basée sur des attributs est donné par [LEH 01] et peut être visualisé à la figure 1. Les nombreux algorithmes de simplification de lignes, basés sur une logique de calculs simple tel que l'algorithme de Douglas-Peucker [DOU 73], sont également rapides et appropriés au contexte de la généralisation à la volée.

[GLO 99] présente un exemple combinant plusieurs algorithmes simples avec une base de données comprenant des attributs précalculés en fonction de l'échelle, en une solution unique pour la généralisation à la volée. Les algorithmes exploitant les hiérarchies sont également appropriés pour des applications contraintes par le temps. Ainsi, l'algorithme par [DOU 73] peut être précalculé et servir de fondement (avec les tolérances de simplification résultantes) à la création d'une structure de données en arbre à l'instar de ce qui a été fait pour l'arbre « binary line generalization » (BLG) décrit par [OOS 95]. [OOS 95] présente également un arbre qui permet une opération de généralisation encore plus complexe, appelé arbre de « generalized area partitioning » (GAP). Cet arbre définit des hiérarchies successives composées d'un agrégat de polygones adjacents au sein d'une mosaïque de polygones.

Le travail effectué par [LEH 01] est dédié à la généralisation cartographique au sens strict, et se concentre sur les processus de généralisation en temps réel pour la visualisation de données spatiales sur différents environnements tels que le Web et les dispositifs mobiles. L'idée repose sur l'affichage *ad hoc* de cartes provenant d'une seule base de données détaillées. Tel que mentionné par [LEH 01], cette approche nécessite des mécanismes de généralisation en temps réel qui sont très

différents du processus traditionnel. Puisque les scénarios possibles incluent également l'utilisation de dispositifs sans-fil (ex. assistants personnels), la carte générée doit également prendre en considération la position de l'utilisateur. Ce qui veut dire que seulement les opérateurs simples et faciles à utiliser peuvent être implantés dans un tel contexte. Dans ce cas précis, seulement les opérateurs de sélection, de simplification de bâtiments et d'agrégation de surfaces ont été implantés et testés sur des données topographiques. Un prototype fonctionnel basé sur le SIG *SmallWorld* avec deux clients différents a été développé. Une première application utilise la librairie de visualisation cartographique en JAVA (appelée *OpenMap*), alors qu'une deuxième exploite le format SVG (voir figure X.1).

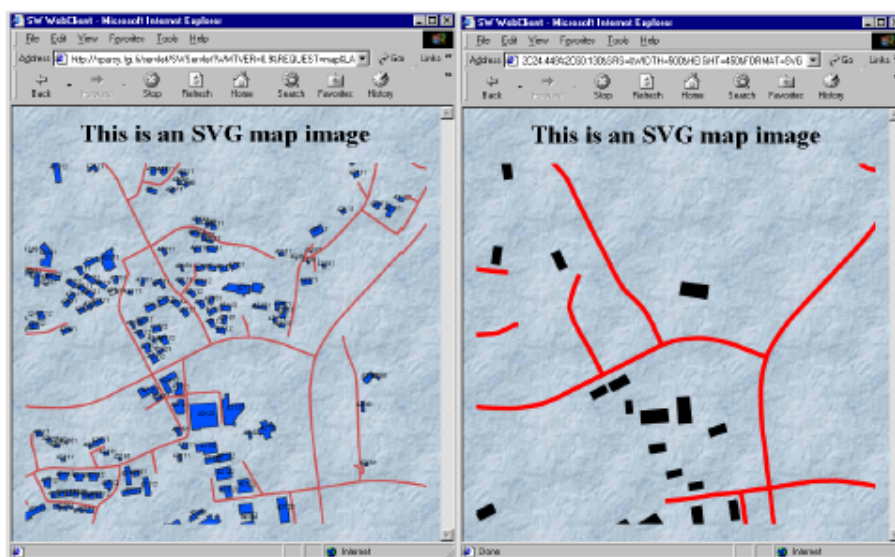


Figure X.1 : Deux affichages SVG d'un même jeu de données, transformé par deux processus différents selon les spécifications de l'utilisateur. Opérateurs de généralisation implantés : sélection et simplification. Source: [LEH 01].

Contrairement à la sélection, à la simplification ainsi qu'à l'agrégation, le déplacement est généralement un opérateur qui ne peut être effectué rapidement (voir le chapitre par Barrault de ce même volume). Une solution possible afin d'accélérer les calculs de déplacements consiste à utiliser l'interpolation de forme (*morphing*) entre les géométries appartenant à deux niveaux d'échelle différents [MON 91], [SED 92]. Toutefois, cette approche requiert une base de données multiéchelle à l'intérieur de laquelle les différentes géométries d'un même objet doivent être interreliées.

12 La généralisation à la volée pour le web

Somme toute, tel qu'il peut être perçu à travers les discussions de ce volume, la généralisation automatique figure toujours au stade de recherche. La généralisation à la volée étant encore plus jeune, des solutions complètes ne peuvent être espérées à l'heure actuelle. Étant donné que de véritables processus à la volée ne sont toujours pas possibles, des approches alternatives basées sur des prétraitements et des bases de données à représentations multiples furent explorées.

1.3.2. État de l'art de l'approche orientée-représentations

L'approche orientée-représentations est basée sur la conservation de plusieurs représentations d'un même objet au sein d'une seule base de données. Cette approche repose essentiellement sur une structuration adéquate de la base de données spatiales.

Dans ce contexte, [KIL 97] fait la distinction entre les termes **base de données multiéchelle** et **base de données à représentations multiples**. Alors que le premier réfère à la présence de plusieurs représentations d'un même objet qui ne varient qu'en fonction de l'échelle, le deuxième terme est employé pour les bases de données constituées de différents niveaux de représentation ayant des degrés d'abstraction différents et qui ne sont pas nécessairement associés à une échelle en particulier. Ainsi, une base de données à représentations multiples peut être utilisée pour différentes applications spatiales.

[JON 00] discute de deux problèmes majeurs liés à la généralisation dynamique de cartes basée sur une base de données multiéchelle. Le premier est la difficulté de concevoir des schémas d'accès aux données spatiales multiéchelles alors que le second provient de la complexité de conception des processus de symbolisation en ligne pour afficher les données et assurer une représentation lisible. Une des limitations de ces schémas multiéchelles est dûe au fait que la représentation est souvent pré-généralisée indépendamment des autres objets, ce qui peut créer des conflits spatiaux lors de changements d'échelle ou de symbologie. [JON 00] propose d'appliquer les techniques de triangulation (la triangulation de Delaunay) pour détecter et résoudre les conflits spatiaux dans un processus en temps réel et ainsi préserver la topologie.

La transmission progressive de données vectorielles est un autre terme qui est associé avec les concepts de généralisation en temps réel et de bases de données multiéchelles. Le premier travail concernant la transmission progressive de données vectorielles sur Internet a été fait par [BUT 99]. Celle-ci affirme que les processus de transmission progressive de données sont inexistants pour les formats vectoriels, alors qu'ils sont bien connus et appliqués avec succès pour les images de format matriciel, où des versions plus grossières sont affichées avant le téléchargement

complet de l'image. Cette technique est essentielle surtout pour les jeux de données volumineux qui doivent être transmis par des liens de communication plutôt lents. Ses recherches ont porté principalement sur la subdivision hiérarchique de données vectorielles en utilisant des structures de données basées sur l'algorithme de Douglas-Peucker [DOU 73], à l'instar du « BLG tree » par [OOS 95].

Un autre travail dans ce domaine a été effectué par [BER 01]. Ces derniers se sont concentrés principalement sur la transmission progressive de données vectorielles en conjonction avec la cartographie à la volée sur le Web. Ils proposent de créer, à partir d'un même jeu de données, des représentations multiples correspondant à différents niveaux de détail. Ceux-ci sont ensuite transmis séquentiellement au client par ordre croissant de niveaux de détail. En fait, seulement les différentiels sont transmis et additionnés aux représentations déjà affichées. Bien que leur solution soit conceptuellement très intéressante, un problème majeur reste à être solutionné en ce qui concerne l'intégration des données transmises du côté du client, non seulement au niveau graphique mais également en terme de relations spatiales afin que la représentation affichée soit correcte et complète à chaque étape.

1.4. Vers de nouvelles solutions combinatoires

Toutes les approches discutées précédemment reposent soit sur des processus de généralisation à la volée, soit sur une base de données multiéchelle. Toutefois, afin de mieux exploiter les avantages des deux approches, il serait logique de développer une solution combinée qui reposerait entre ces deux extrémités.

L'approche proposée par [CEC 01] [CEC 02] est basée sur une telle combinaison entre bases de données multiéchelles et généralisation à la volée. Ceci permet de concevoir des solutions plus flexibles (dû au processus de généralisation à la volée), qui peuvent s'adapter facilement aux besoins des utilisateurs. Également la base de données multiéchelle assure l'efficacité et facilite les opérations complexes de généralisation (ex. *morphing*). La caractéristique principale de cette approche est que selon les spécifications des utilisateurs, les classes d'objets sont sélectionnées à partir du niveau de détail de la base de données qui est le plus approprié à l'échelle de la carte et ensuite raffinées via des algorithmes de généralisation à la volée. Cette approche est basée sur le module GENDEM (figure X.2) qui lui-même se subdivise en sous-modules. Ceux-ci possèdent chacun leurs fonctions spécialisées et interagissent entre eux afin de répondre à la requête de l'utilisateur en obtenant les données nécessaires à partir d'une base de données multiéchelle. Le module client (*Front-end Module*) forme l'interface entre l'utilisateur et le système. Le module de création de cartes comprend trois types d'opération : symbolisation, détection de conflits et opérations de généralisation.

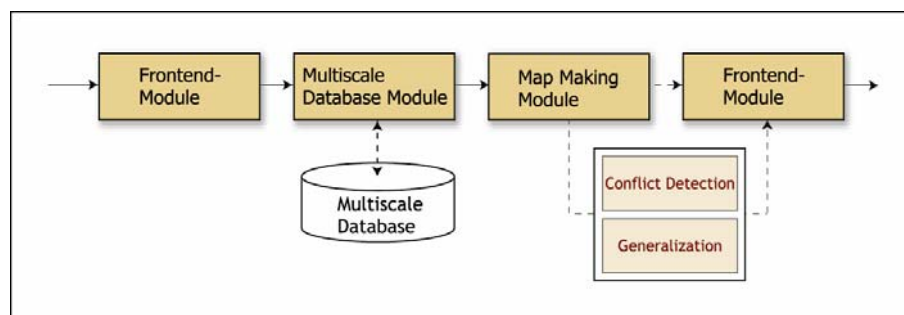


Figure X.2 Le cycle de traitements proposé par [CEC 01] [CEC 02] et les sous-modules qui lui sont associés.

La base de données multiéchelle est utilisée en tant que base pour la génération de cartes et inclus un minimum de deux jeux de données représentant les objets à différents niveaux d'échelle ou de détail, avec des liens de correspondance entre les différents niveaux. Chaque objet ou groupe d'objets doit connaître son homologue à plus grande ou plus petite échelle. Cette information est extrêmement importante et est utilisée afin de simplifier le processus de généralisation. De par ces liens entre les différents niveaux d'échelle, la généralisation peut être vue comme un processus d'interpolation (*morphing*) entre deux géométries différentes.

La connaissance du processus de généralisation cartographique peut être exploitée afin de concevoir le contenu de la base de données multiéchelle, en particulier pour 1) les niveaux de détail, 2) les limites d'applicabilité des niveaux de détail, 3) les opérateurs de généralisation utilisés pour les transformations entre les niveaux de détail. Selon cette approche, la généralisation automatique peut être séparée en deux tâches, ce qui permet de la simplifier. La première tâche s'effectue en prétraitement et consiste à créer la base de données multiéchelle et à enrichir les données. La deuxième tâche s'effectue en ligne et consiste à ajuster les niveaux de détail en utilisant la généralisation à la volée et le *morphing*.

Une autre approche proposée par [BER 02] suggère d'utiliser des patrons de représentations multiples dans le processus même de généralisation automatique. Ces patrons, qui sont en fait des représentations typiques de certains types d'objets, permettent de faciliter certaines opérations de généralisation à la volée afin d'accélérer les temps de réponse et permettre l'analyse en ligne de données spatiales.

Par exemple, différents patrons servant à représenter les bâtiments peuvent être conservés au sein d'une même base de données spatiales. Non seulement ces patrons sont définis pour différents niveaux de détail mais peuvent également partager un même niveau de détail (ex. deux représentations d'un même objet issues de normes

de saisie différentes), ce qui permet l'exploitation de la même base de données dans le cadre d'applications différentes (figure X.3). Par ailleurs, ces patrons s'avèrent particulièrement utiles lors d'une généralisation à la volée. En effet, il est fréquent que certaines formes de bâtiment causent des problèmes ou engendrent des résultats moins satisfaisants lors de leur simplification via des algorithmes de généralisation automatique. Ainsi, il pourrait être avantageux de tout simplement remplacer la géométrie du bâtiment devant être généralisé par une géométrie plus simplifiée (i.e. un patron préalablement enregistré au sein de la base de données) et de la positionner adéquatement.

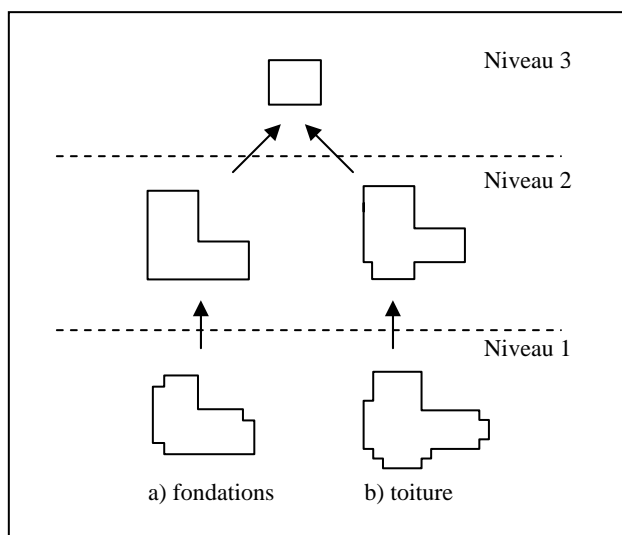


Figure X.3 Deux patrons correspondant à des normes de saisie différentes (et leurs versions simplifiées) servant à représenter un même objet à des niveaux d'abstraction différents et/ou selon des besoins différents.

Un autre exemple concerne les données routières. La généralisation automatique appliquée à des éléments linéaires tels que les réseaux routiers ou hydriques engendre généralement de bons résultats. En effet, étant donné que quatre-vingt pourcent (80%) des objets cartographiques sont linéaires, la généralisation de lignes fut amplement étudiée et par conséquent, les algorithmes concernant ce genre d'objets sont relativement efficaces [MUL 95]. Toutefois, dès que ces réseaux possèdent des éléments plus complexes, la qualité du résultat se détériore rapidement et les temps de traitement sont beaucoup plus élevés. Afin de mieux illustrer ceci, prenons l'exemple d'une route qui se compose d'une partie très

sinueuse. Lors d'une réduction d'échelle, un empâtement des virages est quasi-inévitable ce qui engendre une représentation peu lisible de la route en question (figure X.4).

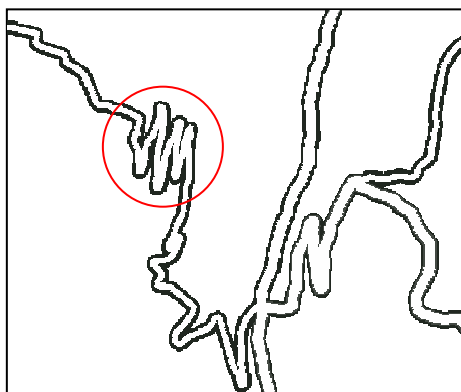


Figure X.4 *Route conflictuelle*⁶

Afin de corriger la situation, il est nécessaire d'appliquer certains opérateurs de généralisation. Toutefois, vu la complexité de ces opérateurs, il est excessivement difficile, voire impossible, d'effectuer cette généralisation rapidement en mode automatique.

En revanche, l'utilisation de patrons de représentations multiples permet de faciliter largement ce processus. Ainsi, au lieu de généraliser la partie de route sinueuse, elle peut tout simplement être remplacée par un patron, préalablement enregistré dans la base de données, correspondant à ce type de géométrie. Évidemment, quelques manipulations (orientation, étirement, etc.) devront être effectuées sur le patron dans le but de conserver la cohérence de la route. Toutefois, il s'agit d'opérations très simples pouvant être effectuées rapidement (figure X.5).

⁶ © BDCarto de l'IGN

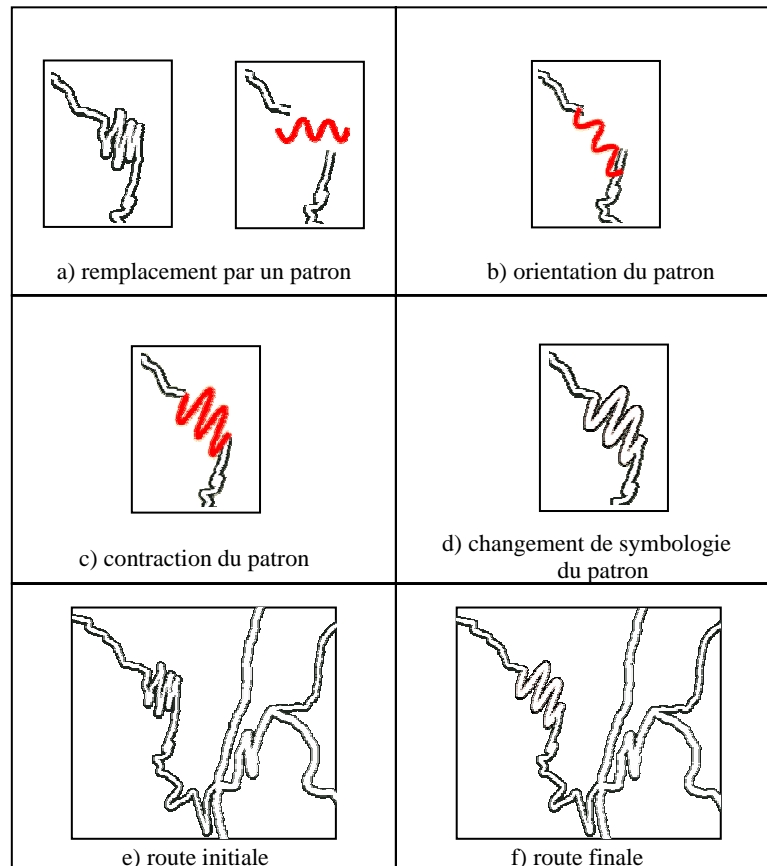


Figure X.5 Utilisation d'un patron de route sinueuse dans le processus de généralisation à la volée

Il doit être mentionné que ce genre de processus nécessite une structuration plus riche au niveau des données. Une structure de données à l'instar de celle actuellement exploitée au ministère des Transports du Québec (MTQ) est à privilégier. Les données sont emmagasinées en tant que « segments atomiques ». Un segment atomique représente la plus petite unité homogène (relativement à l'ensemble de ses propriétés) du réseau routier. Plusieurs critères, tels que le type de pente et le degré de courbure sont utilisés afin de segmenter le réseau routier du Québec. Pour chacun des segments, il est possible de connaître ses coordonnées

d'origine et de destination. Ainsi, via ces informations, il est très facile de localiser un élément potentiellement problématique en vue d'une généralisation automatique et de le remplacer par un patron à la position exacte.

Notons finalement que ces patrons pourraient avoir été définis dès la saisie même des objets en vue de les stocker dans une base de données à représentations multiples. En effet, au moment de la saisie des données, une première numérisation peut correspondre à la géométrie exacte de l'objet alors que des numérisations additionnelles peuvent être effectuées pour l'obtention de géométries simplifiées ou déplacées du même objet. Ces géométries, une fois emmagasinées au sein d'une même base de données, pourront servir à d'éventuels processus de généralisation à la volée.

1.5. Conclusion

Avec l'émergence de la cartographie Web, des dispositifs mobiles ainsi que des outils SOLAP, la généralisation est maintenant sujette à de nouvelles contraintes. Parmi les plus importantes d'entre elles figurent la nécessité de fournir des temps de réponse quasi-instantanés. Étant donné que très peu d'algorithmes de généralisation existants sont aptes à satisfaire un tel besoin, de nouvelles solutions doivent être envisagées afin de permettre une généralisation à la volée.

Jusqu'à tout récemment, les approches proposées reposaient soit sur des processus de généralisation automatique, soit sur une base de données multiéchelle. Toutefois, les limites de chacune des approches apparaissent rapidement. Depuis peu, de nouvelles solutions émergent et proposent de combiner les deux approches en une solution unique pour la généralisation à la volée. Il devient ainsi possible de tirer profit de leurs avantages respectifs tout en palliant à leurs inconvénients. Ces approches suggèrent généralement de jumeler l'exploitation d'une base de données à représentations multiples (possédant différents niveaux de détail où les géométries d'un même objet sont interconnectées) à des processus de généralisation automatique. Ces derniers permettant de générer des produits cartographiques à des échelles arbitraires, non stockées dans la base de données. Un certain compromis est donc réalisé entre les temps de traitement d'une part et le volume de données emmagasinées d'autre part. Par ailleurs, nous croyons qu'une solution combinatoire encore plus formelle peut être envisagée afin d'accélérer et de faciliter les processus de généralisation. Cette solution est basée sur l'exploitation de patrons de représentations multiples dans le cadre du processus même de généralisation automatique. Finalement, il est de notre avis qu'actuellement, l'utilisation indépendante des deux approches ne peut satisfaire les contraintes sous-jacentes à la généralisation à la volée et que le développement de solutions combinatoires est nécessaire à l'obtention de systèmes plus flexibles, répondant mieux aux besoins des

utilisateurs. Ceci dit, du point de vue de ces derniers, peu importe les mécanismes utilisés, il s'agira toujours de généralisation à la volée. L'important pour ceux-ci étant que les données leur soient présentées sans trop de délai!

1.6. Bibliographie

- [ARL 99] METTE A., «Problems in screen map design», *19th International Cartographic Conference Ottawa*, p. 849-857, aout 1999.
- [BÉD 97] BEDARD Y., Spatial OLAP, Vidéoconférence. 2^{ème} Forum annuel sur la R-D, Géomatique VI: Un monde accessible, Montréal, 13-14 novembre 1997.
- [BER 02] BERNIER E., Utilisation de la représentation multiple comme support à la génération de vues de bases de données géospatiales dans un contexte SOLAP, Mémoire de M.Sc, Faculté de foresterie et géomatique, Université Laval, 89 p., 2002
- [BER 01] BERLOTTI M., EGENHOFER M.J., «Progressive Transmission of Vector Map Data over the World Wide Web», *GeoInformatica*, 5(4), p. 345-373, 2001.
- [BUT 99] BUTTENFIELD B.P., «Sharing Vector Geospatial Data on the Internet», *19th International Cartographic Conference Ottawa*, p. 39-45, aout 1999.
- [CEC 01] CECCONI A., WEIBEL R., «Map Generalization for On-demand Mapping», *GIM International*, 15(12), p. 12-15, 2001.
- [CEC 02] CECCONI A., WEIBEL R., BARRAULT M., «Improving Automated Generalization for On-demand Web Mapping by Multiscale Databases», *10th International Symposium on Spatial Data Handling Ottawa*, (submitted), juillet 2002.
- [COD 93] CODD E.F., CODD S.B. & SALLEY C.T., « Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate », *Hyperion white papers*, <http://www.hyperion.com>, 20 pp., 1993.
- [DOU 73] DOUGLAS D., PEUCKER T., « Algorithms for the reduction of the number of points required to represent a digitized line or its caricature», *The Canadian Cartographer*, 10(2), p. 112-122, 1973.
- [ELZ 01] VAN ELZAKKER C., «Use of Maps on the Web» in Kraak, M. J., Brown, A., *Web Cartography: Developments and Prospects*, Taylor and Francis, p. 21-36, 2001.
- [FER 01] FERLINGA W., «File formats and plugins» in Kraak, M. J., Brown, A., *Web Cartography: Developments and Prospects*, Taylor and Francis, p. 177-193, 2001.
- [GLO 99] GLOVER E., MACKANESS, W.A., «Dynamic generalisation from a single detailed database to support web based interaction», *19th International Cartographic Conference Ottawa*, p. 1175-1184, aout 1999.
- [GOW 97] GOWER R., PEPPER J. & EDWARDS T., «Landplan - Automated Generalisation comes of Age», 18th International Cartographic Conference Stockholm, p. 126-133, June 1997.
- [HAN 98] HAN J., STEFANOVIC N. & KOPERSKI K., Selective materialization: An Efficient Method for Spatial Data Cube Construction, *Proceedings of the 1998 Pacific-Asia*

Conference on Knowledge Discovery and Data Mining (PAKDD'98), Melbourne, Australia, pp. 144-158, 1998.

- [JON 00] JONES C. B., ABDELMOTY A.I., LONERGAN M.E., VAN DER POORTEN P., ZHOU S., «Multi-scale spatial database design for online generalisation», *9th International Symposium on Spatial Data Handling Beijing*, p. 7b.34-7b.44, aout 2000.
- [KIL 97] KILPELAINEN T., *Multiple Representation and Generalization of Geo-Databases for topographic Maps*, PhD Thesis, Finnish Geodetic Institute, University of Helsinki, 1997.
- [KRA01] KRAAK M. J., BROWN A., *Web Cartography: Developments and Prospects*, Taylor and Francis, London, 2001.
- [LEH 01] LEHTO L., KILPELÄINEN T., «Generalizing XML-Encoded Spatial Data on the Web», *20th International Cartographic Conference Beijing*, p. 2390-2396, aout 2001.
- [MAR 01] MARCHAND P., BÉDARD Y. ET EDWARDS G., 2001, « A hypercube-based method for spatio-temporal exploration and analysis ». *GeoInformatica*, soumis mai 2001.
- [MON 91] MONMONIER, M., «The role of interpolation in feature displacement», Buttenfield B.P. & McMaster R.B., *Map generalization: Making rules for knowledge representation*, Longman, Harlow, p. 189-203, 1991.
- [MUL 95] MÜLLER J.C., WEIBEL R., LAGRANGE J.P. & SALGÉ F., «Generalization : state of the art and issues» , dans Müller, J.C., J.P., Lagrange & R., Weibel, (éditeurs), *GIS and Generalization: Methodology and Practice*, Taylor & Francis, Londres, pp. 3-17, 1995.
- [OOS 95] VAN OOSTEROM P., SCHENKELAARS V., «The development of an interactive multi-scale GIS», *International Journal of Geographical Information Systems*, 9, p. 489-508, 1995.
- [PAP 01] PAPADIAS D., KALNIS P., ZHANG J. & TAO Y., « Efficient OLAP Operations in Spatial Data Warehouses », *Technical Report HKUST-CS01-01*, 2001.
- [PET 99] PETERSON M.P., Trends in Internet map use, A second look. In Touch the Past, Visualize the Future, *Proceedings 19th International Cartographic Conference*, Ottawa, Canada, edited by Keller, C.P., Section 5 : Capitalizing on new technologies (Ottawa : Organizing Committee for Ottawa ICA 1999) CD-ROM, 1999.
- [RIV 01] RIVEST S., BÉDARD Y. & MARCHAND P., Towards better support for spatial decision-making: Defining the characteristics of Spatial On-Line Analytical Processing (SOLAP), *Geomatica, the journal of the Canadian Institute of Geomatics*, Vol. 55, no. 4, pp. 539-555, 2001.
- [RUS 90] RUSAK MASUR E, CASTNER H.W., «Horton's ordering scheme and the generalisation of river networks», *The Cartographic Journal*, 27(2), p. 104-112, 1990.
- [SED 92] SEDERBERG T.W., GREENWOOD E., «A physically based approach to 2-D shape blending», *Computer Graphics*, 26(2), p. 25-34, 1992.
- [STE 00] STEFANOVIC N., HAN J. AND KOPERSKI K., « [Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes](#) », *IEEE Transactions on Knowledge and Data Engineering*, 12(6): 938-958, 2000.

- [THO 99] THOMSEN E., SPOFFORD G., CHASE D., *Microsoft OLAP Solutions*, John Wiley & Sons, 509 p., 1999.
- [ZHO 99] ZHOU X., TRUFFET D. & HAN J., « Efficient Polygon Amalgamation Methods for Spatial OLAP and Spatial Data Mining », *Proceedings of the 6th International Symposium on Large Spatial Databases (SSD'99)*, Hong Kong, pp. 167-187., 1999.