



Intégration Attributs-Représentation dans les systèmes Cartographiques. Formalisme pour définir un espace intermédiaire hétérogène. Généralisation aux représentations composites et non simplement connexes

Jean-Christophe Lecoq, Michel Mainguenaud

► To cite this version:

Jean-Christophe Lecoq, Michel Mainguenaud. Intégration Attributs-Représentation dans les systèmes Cartographiques. Formalisme pour définir un espace intermédiaire hétérogène. Généralisation aux représentations composites et non simplement connexes. Jun 2004, 2004. <sic_00001238>

HAL Id: sic_00001238

https://archivesic.ccsd.cnrs.fr/sic_00001238

Submitted on 7 Dec 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Intégration Attributs-Représentation dans les systèmes Cartographiques

Formalisme pour définir un espace intermédiaire hétérogène. Généralisation aux représentations composites et non simplement connexes

Jean-Christophe Lecoq — Michel Mainguenaud

Laboratoire Perception Système Information (PSI), FRE CNRS 2645

INSA de Rouen, 1060 Avenue de l'Université

76801 St Etienne du Rouvray Cedex - France

{jean-christophe.lecoq,michel.mainguenaud}@insa-rouen.fr

RÉSUMÉ. L'intégration de composantes hétérogènes est un sujet actif de la communauté de la modélisation conceptuelle des systèmes d'information. Les systèmes à représentation cartographique n'échappent pas à ce besoin tant l'hétérogénéité entre les composantes des entités manipulées est grande. Aux représentations sont en effet associés des attributs alphanumériques, et l'usage simultané de ces deux composantes n'est pas évidente.

Nous présentons la généralisation d'une technique d'intégration par sélection de sémantique. La généralisation est appliquée à des représentations composites et non simplement connexes afin de déterminer un espace intermédiaire de sélection.

ABSTRACT. The integration between heterogeneous components is an active field of research for the conceptual database modelling community. The spatial representation systems manage data having a great level of heterogeneity. Indeed, with representation are associated alphanumeric attributes. The use of both components at the same time is not easy.

We present a generalization technique defined by a semantics selection process. The generalization is applied to composite and non simply connected representations. By this way, we define a middle space for selection.

MOTS-CLÉS: Sélection de sémantique, Intégration hétérogène, Représentations composites et non simplement connexes

KEYWORDS: Semantics selection, Heterogeneous integration, Composite and non simply connected representations

Introduction

Les systèmes d'information cartographique décomposent logiquement les représentations en une palette de modèles. Cette palette s'étend d'une faible structuration (photographie) à une représentation sémantiquement riche (représentation spatiale). Chaque représentation, aussi structurée soit elle, n'est pas une information isolée et suffisante du système car par nature incomplète. Une autre catégorie d'information souvent négligée est aussi disponible : aux représentations sont en effet associés des attributs alphanumériques, qui bien que de composition hétérogène par rapport à la représentation, présentent un complément d'information particulièrement riche.

Les attributs alphanumériques varient dans leur sémantique et dans l'usage. Pour les modèles d'image de bas niveau, les attributs comblent le *fossé de capteurs* séparant la réalité de la représentation [AIE 02]. Pour les représentations spatiales plus structurées, les attributs sont factuels et décrivent des propriétés non visuelles des entités. Au final, la structure logique est donc **rigoureusement identique**, constituée d'une agrégation de composantes hétérogènes entre attributs alphanumériques (procéduraux ou factuels) et représentations. En conséquence de cette structure, une question présentant une réelle problématique pour l'intégration du système d'information est de comprendre la dépendance entre ces composantes hétérogènes.

Dans cet article, nous améliorons une technique d'intégration hétérogène attribut - représentation. Cette amélioration généralise l'approche de [MAI 94, CLA 99] sur la fermeture des opérateurs. Elle concerne la définition d'une technique pour sortir la méthode d'une approche conceptuelle. Nous tendons vers une méthode dirigée par la représentation, ce qui nous permet de l'appliquer à des représentations différentes.

Dans la section 1, nous présentons la technique de fermeture ainsi que la généralisation. Dans la section 2, nous présentons un exemple conceptuel de niveau schéma pour valider la méthode. Cet exemple concerne la généralisation à des représentations composites et non simplement connexes. Nous terminons par une conclusion.

1. Fermeture et Généralisation

Nous nous intéressons à la façon de définir des opérateurs globaux à la structure hétérogène [alphanumérique, représentation]. Cette technique est la fermeture des opérateurs, et une implémentation est exposée dans [MAI 94]. Son usage reste confidentiel dans la littérature.

Nous exposons deux aspects de la fermeture. Dans une première partie en 1.1, nous présentons les concepts utilisés pour la fermeture et en particulier le mécanisme de sélection de sémantique. Dans une seconde partie en 1.2, nous illustrons ce modèle à l'aide d'un exemple. Nous terminons cette section dans la partie 1.3 où nous exposons notre proposition de généralisation.

1.1. Introduction à la fermeture de [MAI 94, CLA 99]

Le problème adressé par la fermeture est le suivant : comment à partir d'un opérateur sur une représentation obtenir un opérateur sur l'ensemble des composantes d'une agrégation de composantes hétérogènes ? La réponse se trouve dans les algèbres hétérogènes en considérant un opérateur global comme composé d'un ensemble d'opérateurs locaux de sémantique "homogène" par rapport à celle de l'opérateur global. Dans le cas réel, la fermeture est asymétrique puisque l'opérateur spatial existe déjà et doit être globalisé à l'agrégat. Le problème se réduit alors à définir un opérateur alphanumérique associé. Ce dernier équivaut à choisir la sémantique pertinente suivant le prédicat spatial donné, ce qui représente une opération de sélection d'attributs au niveau schéma.

Le schéma de la partie alphanumérique du système présente par rapport à un schéma classique un double niveau de structuration. En effet, à un schéma classique d'attributs factuels sont ajoutées les instances de deux concepts : la Topologie et la Granularité [BAR 81]. Ces concepts sont nécessaires à la sélection des attributs. La Topologie précise si un attribut est valide sur la limite (L), l'intérieur (I) ou la globalité (G) de la représentation spatiale. La Granularité indique si un attribut est valide sur un sous-ensemble (SE) ou l'intégralité (INTG) de cette représentation. Chaque attribut de la base, représentant la partie factuelle d'une "entité géographique" (e.g., une Ville ou une Forêt) est classé suivant ces concepts. Nous définissons cette association par le terme de *classification*.

L'opérateur alphanumérique associé (pour la sélection de sémantique) se construit par une étape de sélection intentionnelle, c'est-à-dire de niveau schéma. La sélection intentionnelle a pour but de constituer la partie intentionnelle de la base. C'est une projection au sens relationnel [FLO 96] qui utilise les valeurs des concepts Topologie / Granularité pour déterminer les attributs pertinents. À chaque opérateur global (à partir de l'opérateur spatial) est défini un tableau permettant de déterminer les attributs qui seront sélectionnés. Au final, le tableau se présente comme une fonction f de l'ensemble (Topologie \times Granularité \times Opérateur) vers l'ensemble ((Topologie \times Granularité) \cup \neg Pertinent), avec \neg Pertinent représentant le terme spécifiant la non pertinence de l'attribut.

Cette approche permet de transférer la sémantique de l'attribut dans un *espace intermédiaire* (Topologie \times Granularité) pour déterminer l'opérateur global. À chaque attribut pris dans l'ensemble fini des attributs du système, correspond un point unique dans cet espace. La définition de ce point est la classification, et représente une application injective dont le but est de définir la pertinence d'un attribut suivant une représentation de l'espace. Elle se définit comme une forme d'unification de schéma puisque le schéma spatial est ainsi partiellement intégré au schéma alphanumérique.

1.2. Illustration

Pour illustrer ces concepts d'intégration, nous présentons une base Jouet. Nous ne nous intéressons qu'à la partie alphanumérique de la base. Celle-ci se compose de l'ensemble d'entités {Ville, Forêt}. Le tableau 1 expose les attributs ainsi que les classifications associées. La classification statique des attributs se définit à partir de justifications géographiques, physiques ou administratives. Par exemple, le Maire d'une Ville est un attribut spatialement classé comme (Topologie : global, Granularité : sous-ensemble) car pour n'importe quel sous-ensemble de la représentation spatiale d'une ville, aussi bien pour la limite que l'intérieur, il est pertinent.

Attribut	Signification	Topo	Gran
Entité Ville			
Maire	Maire de la Ville	G	SE
Population	Population de la ville	I	INTG
MurExt	Type de la limite extérieure	L	SE
Entité Forêt			
Essence	Classe des essences	I	SE
Hauteur	Hauteur maximale	I	INTG
Observateur	Employé effectuant des études à l'intérieur	I	SE
GardeBarrière	Employé contrôlant la frontière	L	SE
SommePérimètre	Somme des frontières	L	INTG

Tableau 1. *Table des attributs de la base*

Nous présentons un exemple avec l'opérateur d'intersection (\cap). Le tableau 2 définit les règles de sélection intentionnelle pour l'opérateur (commutatif) d'Intersection. Comme exemple de résultats, nous exposons dans le tableau 3 le schéma de l'intersection entre la Ville et la Forêt menant à la nouvelle entité (Ville \cap Forêt). La nouvelle entité est hybride puisqu'elle contient des attributs provenant des deux entités parentes. Pour cet opérateur, seuls les attributs qualitatifs comme les noms ont été gardés. Par contre la Population ou la Hauteur n'ont pas été retenues.

Granularité Topologie	intégralité	sous-ensemble
limite	\neg Pertinent	(L, INTG)
intérieur	\neg Pertinent	(I, SE)
global	\neg Pertinent	(G, SE)

Tableau 2. *Règles de sélection intentionnelle de l'intersection (\cap), commutatif.*

Attribut	Topologie	Granularité
Ville.Maire	G	SE
Ville.MurExt	L	INTG
Forêt.Essence	I	SE
Forêt.Observateur	I	SE
Forêt.GardeBarrière	L	INTG

Tableau 3. Schéma du tableau résultant de $(Ville \cap Forêt)$

1.3. Généralisation

Nous sommes intéressés par la généralisation des concepts de Topologie et de Granularité. Ces concepts sont en effet issus d'une analyse géographique et sont donc par conséquent bornés à leur cadre de définition. En particulier, deux restrictions se font ressentir :

- 1) La sémantique associée à une représentation plus spécifique n'est pas supportée (e.g., représentation composite, non simplement connexe),
- 2) Toute représentation dans un cadre de représentation différent n'est pas supportée (e.g., image).

Nous souhaitons retrouver les concepts à partir de représentations réelles. Cela nécessite la définition d'un espace intermédiaire "variable" d'intégration de schémas. L'intégration de schémas doit donc être assurée même si les propriétés de représentation ne sont pas clairement identifiées. C'est une approche ascendante des représentations réelles vers les concepts qui pose ouvertement des problèmes de généralisation. Elle permet de générer dynamiquement un espace intermédiaire d'intégration sans connaître la représentation conceptuelle impliquée. En ce sens, cela représente une approche bénéfique pour le domaine multimédia en général où les concepts pour toutes les représentations ne sont pas disponibles. Notre approche est donc quantitative à défaut d'être qualitative. Elle se fonde sur trois postulats :

- 1) Nous supposons que les propriétés du type sont équivalentes aux propriétés de la représentation, cette équivalence nous permet de nous affranchir partiellement de l'aspect sémantique.
- 2) Nous supposons que la construction du type de la représentation contient implicitement les connaissances sur les concepts de la représentation. Une approche analytique à l'aide de recherche de similitude doit permettre de retrouver les concepts de la nouvelle représentation.
- 3) Nous ne partons pas de rien mais d'une représentation de base dont les concepts sont connus pour l'apprentissage [MAI 94]. Notre démarche est de type différentielle : à une petite variation δ du type de la représentation correspond une autre variation Δ des concepts. C'est la loi de variation $\Delta(\delta)$ que nous cherchons.

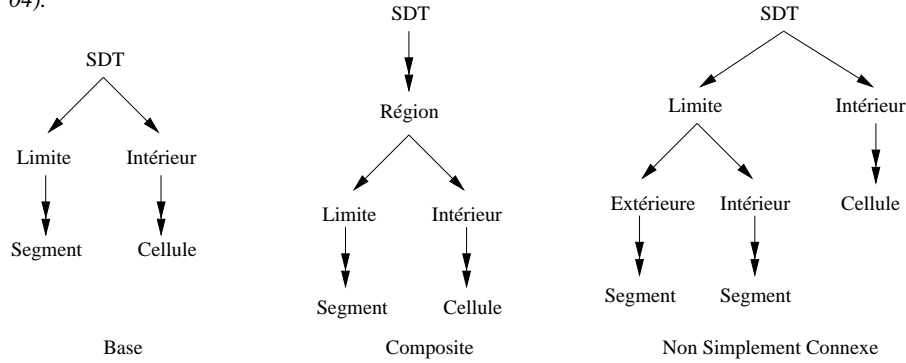


Figure 1. *Grappe des constructeurs pour les représentations spatiales*

La méthode de généralisation que nous utilisons consiste à exploiter analytiquement les propriétés locales des constructeurs de base, et de les combiner hiérarchiquement avec des règles de combinaison pour former un concept plus global. Le constructeur d'ensemble représente la propriété de Granularité alors que le constructeur d'agrégation représente celui de Topologie. La technique de combinaison se fait sous la forme d'un graphe à partir d'une transformation du type de la représentation. La classification d'un attribut est associée à un noeud de ce graphe dual, c'est-à-dire sur tout un chemin menant de la racine jusqu'au noeud ciblé.

Pour simplifier la présentation, nous ne présentons pas les règles de construction du graphe dual.

2. Applications aux représentations spatiales évoluées

Dans cette section, nous présentons des applications spatiales pour valider les concepts de généralisation, en mesurant l'influence sémantique sur la base de données Jouet. Les applications concernent l'extension des concepts de fermeture au cas de représentations spatiales évoluées : les représentations composites dans la partie 2.1 et non simplement connexes dans la partie 2.2. Nous présentons les évolutions de structure des types de représentation (SDT) dans la figure 1. Le constructeur d'ensemble est modélisé en flèche double alors que l'agrégation est en flèche simple.

2.1. Représentation Composite

Nous définissons la représentation composite par l'existence d'un ensemble de zones autonomes disjointes deux à deux. Nous supposons que ces zones sont sim-

plement connexes (pas de trous dans la représentation). La représentation spatiale se décompose alors en deux niveaux, le niveau *Composite* (la représentation) et le niveau *Région* (un des éléments de la représentation). Le problème est de déterminer les propriétés et les informations de niveau Composite puisque ces dernières sont liées par un constructeur d'ensemble au niveau plus élevé.

À partir de la méthode de généralisation, le concept associé à la représentation est composé par la chaîne conceptuelle (Granularité, Topologie, Granularité). Pour différencier les concepts de Granularité entre eux, nous nommons les concepts par rapport à leurs niveaux de structuration du type composite. Soit GranIntra (Granularité Intra Région) le concept de Granularité associé au niveau Région (sous-ensemble ou intégralité pour les régions considérées). Soit GranInter (Granularité Inter Région) le nouveau concept défini au niveau SDT (sous-ensemble ou intégralité de toutes les régions), voir figure 1. Nous résumons les nouvelles classifications des attributs dans le tableau 4. Nous observons que la classification de la Topologie ne change pas par rapport à la classification Jouet initiale en tableau 1. Cette évolution de classification est donc à Topologie constante.

Attribut	Topologie	GranInter	GranIntra
Entité Ville			
Maire	G	SE	SE
Population	I	INTG	INTG
MurExt	L	SE	SE
Entité Forêt			
Essence	I	SE	SE
Hauteur	I	SE	INTG
Observateur	I	INTG	SE
GardeBarrière	L	SE	SE
SP	L	INTG	INTG

Tableau 4. *Nouvelles classifications des attributs alphanumériques*

Avec le nouvel espace intermédiaire, l'opérateur de sélection intentionnelle doit être effectué sur un cube. L'usage d'un cube implique une cardinalité plus grande. Cependant nous pouvons supposer raisonnablement que ce cube est rempli de valeurs nulles pour des plans entiers, ce qui simplifie le processus.

Nous souhaitons définir l'opérateur alphanumérique équivalent au prédicat spatial correspondant à l'intersection entre la Ville et Forêt. Soit \cap^3 le nouvel opérateur d'intersection, et \cap^2 l'ancien. La Topologie n'étant pas un critère déterminant pour l'intersection, le problème se résume au choix des catégories sémantiques pertinentes pour les Granularités. Du fait que l'intersection ne prenne qu'un sous-ensemble de la représentation, une seule catégorie de sémantique est conservée : (GranInter : sous-ensemble, GranIntra : sous-ensemble). La sélection appliquée par les Granularités est ensuite affinée par la Topologie. Dans le cas indéfini d'une Topologie en limite, nous classons tous

^e soumission à *Modèles et Applications pour le Document Cartographique Numérique (SDN 04)*.

les attributs en "local" pour minimiser les risques. Cela signifie que la classification des attributs est associée à chaque région, qui deviennent donc structurellement multivalués. Le tableau 5 représente les règles de sélection / reclassification des attributs pour \cap^3 . Le tableau 6 expose ensuite les sélections des attributs suivant les espaces intermédiaires des opérateurs d'intersection. Nous remarquons que l'attribut Observateur n'est pas sélectionné par l'opérateur \cap^3 . Cela nous permet de vérifier que la spécification plus importante des concepts par rapport à la représentation (i.e., espace intermédiaire plus complexe) restreint la sélection.

Topologie \ Granularité	sous-ensemble
limite	(L, SE, INTG)
intérieur	(I, SE, SE)
global	(G, SE, SE)

Tableau 5. Règles de sélection intentionnelle de l'intersection \cap^3

Attribut	\cap^2	\cap^3
Maire	(G, SE)	(G, SE, SE)
Population	\neg Pertinent	\neg Pertinent
MurExt	(L, INTG)	(L, SE, INTG)
Essence	(I, SE)	(I, SE, SE)
Hauteur	\neg Pertinent	\neg Pertinent
Observateur	(I, SE)	\neg Pertinent
GardeBarrière	(L, INTG)	(L, SE, INTG)
SP	\neg Pertinent	\neg Pertinent

Tableau 6. Sélection des attributs suivant les opérateurs d'intersection

2.2. Représentation Non Simplement Connexe

Les régions non simplement connexes se caractérisent par le fait que pour aller d'un point à un autre dans la représentation, tous les chemins possibles continus ne sont pas utilisables. D'un point de vue structurel, la représentation définit deux limites. Ces deux limites sont d'abord la *limite extérieure* qui est identique à celle de la représentation classique. Nous définissons ensuite la *limite intérieure* pour borner les ensembles de points interdits. Dans le but de simplifier l'approche, cette représentation n'est pas récursive puisque aucune référence n'est faite (située sur l'intérieur) pour une autre représentation.

À partir du type dual, les modifications sur les concepts pour l'espace intermédiaire sont les suivantes :

1) Le concept de Topologie est augmenté avec la prise en compte de la limite intérieure. Le nouveau concept est donc obtenu par énumération combinatoire : <intérieur (I), limite extérieure (LE), limite intérieure (LI), I+LE, I+LI, ..., global (G)>.

2) Le concept classique de Granularité sur le premier constructeur d'ensemble est inchangé.

3) Un nouveau concept optionnel apparaît en charge de la gestion des limites intérieures. Soit GranLimInt (Granularité Limite Intérieure) ce nouveau concept.

Le dernier concept n'est pas valide pour tout cheminement dans l'espace intermédiaire. Du fait de l'asymétrie du type, il admet une valeur supplémentaire d'inactivité de manière à inhiber son influence : GranLimInt = {SE, INTG, **inactif**}. La chaîne que nous nommons ainsi reprend les deux concepts classiques Topologie et Granularité et est complétée avec le nouveau concept GranLimInt. Nous re-classons les attributs dans le tableau 7.

Attribut	Topologie	Granularité	GranLimInt
Entité Ville			
Maire	G	SE	SE
Population	I	INTG	inactif
MurExt	LE	SE	inactif
Entité Forêt			
Essence	I	SE	inactif
Hauteur	I	INTG	inactif
Observateur	G	SE	SE
GardeBarrière	LI	SE	SE
SP	I+LI	INTG	INTG

Tableau 7. *Nouvelles classifications des attributs alphanumériques*

Comme application, nous définissons "l'influence" de la Forêt sur la Ville avec un opérateur évolué dérivé du Collage Spatial. Cet opérateur se définit spatialement comme le collage des limites extérieures / intérieures des régions, de manière à "engluer" un élément dans un autre. Topologiquement, cet opérateur de collage (\diamond) correspond à la fermeture de l'élément qui est inclus (a), sur la limite intérieure de celui qui reçoit (b) et se définit par l'opération binaire : $a \diamond b = a \cup (a \triangleleft \sigma^- b)$, avec $\sigma^- b$ l'adjacence concave de b (soustraction de l'ensemble avec ses trous laissant les limites intérieures). Comparé à un opérateur d'adjacence classique [MAG 91], cet opérateur de collage prend en compte alphanumériquement l'intégralité de la limite.

Du point de vue alphanumérique, la définition de l'opérateur associé se traduit dans le tableau de règles des sélections d'attributs. Quelques éléments justifient les choix de sélection :

- Pour l'élément incluant, seuls les attributs sur la limite intérieure sont concernés,

^e soumission à *Modèles et Applications pour le Document Cartographique Numérique (SDN 04)*.

- La sélection sur la limite intérieure ne se fait que pour les attributs associés sur un sous-ensemble, car au plus toutes les régions intérieures sont concernées,
- Un transfert de sémantique a lieu entre les attributs classés en limite intérieure, qui deviennent associés à la limite extérieure avec GranLimInt inactif.

Comme exemple de résultats, nous présentons dans le tableau 8 les attributs sélectionnés pour la nouvelle entité (Forêt \diamond Ville). Nous ne pouvons pas comparer les résultats de cet opérateur avec l'espace intermédiaire classique, car cet opérateur ne peut s'appliquer dans ce cas. Les nouveaux concepts permettent donc d'exprimer un opérateur associé à une représentation plus fine.

Attribut	Topologie	Granularité	GranLimInt
Ville.Maire	G	SE	SE
Ville.Population	I	INTG	inactif
Ville.MurExt	LE	SE	inactif
Forêt.Observateur	LE	SE	inactif
Forêt.GardeBarrière	LE	SE	inactif

Tableau 8. *Résultat de (Forêt \diamond Ville)*

Conclusions

Nous avons présenté dans cet article une généralisation d'une technique d'intégration hétérogène. Cette technique se résume en une sélection de sémantique. Notre contribution s'est orientée sur la généralisation de la méthode de manière à prendre en compte des représentations réelles. Pour généraliser à tout cadre sémantique, l'approche que nous définissons se présente comme une généralisation ascendante des concepts de sélection à partir de la représentation. Nous avons décidé d'orienter le choix de la méthode en plaçant le focus sur l'organisation des constructeurs de base du langage. La méthode est donc analytique

Pour tester la validité de cette méthode, nous avons exposé une application spatiale. Cela se caractérise par la prise en compte de propriétés de représentation plus évoluées comme les régions composites et non simplement connexes. Dans le premier cas, nous illustrons le fait qu'à sémantique constante l'opérateur alphanumérique de sélection associé devient plus stricte. Dans le second cas, nous pouvons définir la fermeture sur un opérateur de sémantique plus évoluée en rapport avec la représentation. Ces deux évolutions soulignent une augmentation du pouvoir d'expression global. Mais en conséquence, l'approche devient dépendante de la représentation.

3. Bibliographie

- [AIE 02] AIELLO M., MONZ C., TODORAN L., WORRING M., « Document understanding for a broad class of documents », *International Journal on Document Analysis and Recognition IJDAR*, vol. 5, n° 1, 2002, p. 1-16.
- [BAR 81] BARRERA R., BUCHMANN A., « Schema definition and query language for geographical database system », *IEEE Transactions on Computer Architecture : Pattern Analysis and image Database Management*, vol. 11, 1981, p. 250-256.
- [CLA 99] CLARAMUNT C., MAINGUENAUD M., « A revisited database projection operator for network facilities in a GIS », *Informatica*, vol. 23, 1999, p. 187-201.
- [FLO 96] FLORY A., LAFOREST F., *Les bases de données relationnelles*, Economica, 1996.
- [MAG 91] MAGUIRE D., GOODCHILD M., RHIND D., *Geographical information systems - Volume II - Applications*, John Wiley and Sons, 1991.
- [MAI 94] MAINGUENAUD M., « Consistency of Spatial Database Query Results », *Computer Environment and Urban Systems*, vol. 18, n° 5, 1994, p. 333-342, Pergamon, Elsevier.