

L'interrogation des bases de données champs continus basée sur une approche figurative

Fatiha Guerroudji Meddah*, Hafida Belbachir*, Robert Laurini**

*Laboratoire LSSD, USTOMB Oran

fguerroudji@yahoo.fr

h_belbach@yahoo.fr

** Laboratoire LIRIS Insa de Lyon 69621

– Villeurbane Cedex – France

Robert.Laurini@insa-lyon.fr

Résumé. Dans les sciences de l'information géographique les champs continus sont utilisés pour modéliser et représenter des phénomènes continus naturels et environnementaux. Ils sont d'un grand intérêt pour décrire la distribution de plusieurs propriétés physiques qui varient continuellement dans l'espace et dans le temps, comme la météorologie. Notre article introduit dans les sections suivantes les notions afférentes à l'information géographique de type champ et présente les spécifications fonctionnelles de notre environnement visuel qui permet aux utilisateurs manipulant ce type de données de formuler leurs requêtes d'une manière intuitive et simplifiée.

Mots clés: Base de données géographiques, champ continu, champ scalaire, champ vectoriel, requête visuelle.

1 Introduction

L'interrogation est sans contexte l'opération la plus importante dans la vie d'une base de données géographique. Cependant, l'accès à l'information géographique et l'appropriation de celle-ci par ses différents utilisateurs, en particulier les non experts demeure une chose pas facile. En effet, vu la rigidité des langages d'accès mis à sa disposition et sa méconnaissance de l'organisation des informations au sein de la base de données, l'utilisateur se trouve souvent devant une information disponible mais non exploitable. La mise en place de langages de récupération de données très pertinents qui intègrent des composantes spatiales et temporelles selon une approche spécifique en termes d'interface, d'interrogation et de visualisation se veut donc un impératif. Notre travail décrit dans ce papier s'inscrit dans ce contexte et aborde la problématique du dialogue pour l'accès à l'information géographique de type champ continu.

1.1 Préliminaires

De nombreux auteurs ont donné des définitions et des descriptions multiples du terme champ [16] [9], [15], [13]. Plus généralement, et c'est la définition que nous

adoptons, un champ est une portion de l'espace où la force appliquée en un point dépend de sa position seule. Il est défini formellement par:

$$CF=f(g) \quad (1)$$

Où g est une région de l'espace où s'applique une force définie par f .

Ces champs dans le domaine de l'information géographique sont appelés «Continuous Field » ou champs continus appellation qui renforce la notion de continuité des valeurs des champs. Le nombre de points contenu dans un champ est infini, la représentation des valeurs de ce champ est donc continue.

La conséquence est que l'on devrait stocker une infinité de valeurs; mais dans la pratique seul un nombre limité de valeurs est stocké, impliquant l'existence de procédures puissantes d'interpolation afin d'estimer les valeurs demandées.

Cette remarque implique l'existence de structures particulières de bases de données dites : bases de données de type champ ou base de données champ continu [9] [15].

Des exemples typiques des champs continus dans un espace à 3 dimensions sont la distribution de la température ou de la pression sur un territoire.

Les Champs continus se distinguent selon deux types : les champs scalaires où une simple valeur est définie en un point (ex. la température) et les champs vectoriels où des valeurs multiples sont assignées en un point (ex le vent contenant deux champs scalaires, comme la vitesse et la direction).

La section suivante du présent article dresse un état de l'art sur la problématique de l'interrogation des bases de données géographiques en général et les bases de données champs continus en particulier.

2 Problématique de l'interrogation des bases de données géographiques

La formalisation d'un langage d'interrogation des bases de données discrètes ou continues dans les sciences de l'information géographique est un thème de recherche important des vingt dernières années.

A l'heure actuelle, deux problématiques majeures sont examinées par les chercheurs dans ce domaine : une première catégorie tente de répondre à la question Quels sont les types de requêtes géographiques généralement spécifiées par les utilisateurs ? cela conduit à établir non seulement une classification grossière des types de requêtes géographiques mais aussi à s'interroger sur les types de requêtes les plus fréquemment formulées par une catégorie d'utilisateurs donnée.

Une deuxième catégorie se donne pour objectif de répondre à la question Quel mode d'interrogation (et d'interaction) préconiser pour un type de requête géographique donné ?

Cette question est analysée dans l'objectif de s'assurer que le mode d'interrogation proposé est celui le plus naturel pour les utilisateurs. Il est aussi question de prendre en compte les préférences de ces utilisateurs et de proposer une interface s'adaptant à leurs profils.

Dans cette deuxième catégorie de recherche, plusieurs travaux ont vu le jour lors de ces dernières années, basés sur diverses approches :

La première approche utilise directement les principes d'algèbre et de géométrie.

Les travaux de Tomlin [21] [22] basés sur l'algèbre cartographique sont les plus connus.

Une seconde approche est celle basée sur l'usage des langages formels souvent issus du standard SQL, avec les travaux d'Egenhofer [4] sur les extensions du SQL, le PSQL [18], le SQL3 [19] et bien d'autres variantes.

Dans cette approche, on extrait de l'information, à partir du système, à l'aide de requêtes construites. Cette approche est par nature non adaptée aux utilisateurs naïfs (ne maîtrisant pas l'outil informatique) car elle exige un apprentissage souvent laborieux.

La troisième approche en recherche dans la communauté des SIG est celle où l'accès à l'information géographique se fait par l'usage exclusif de la langue naturelle. Cette approche tourne autour du langage naturel dans le contexte des expressions spatiales, appuyée fortement par la recherche en cognition, en langage et en informatique.

Or, les phénomènes propres à la langue naturelle (paraphrase, inférence, polysémie etc.) sont très difficiles à maîtriser dans une perspective de compréhension, cette approche se heurte actuellement aux problèmes posés généralement par le traitement ayant pour objectif la compréhension de la langue naturelle. En effet, la formulation des requêtes peut s'avérer longue (un dessin vaut mille mots).

De plus les requêtes verbales sont encore difficiles à interpréter quoique des solutions récentes soient proposées dans [17].

La quatrième approche est celle qualifiée de figurative. Les interfaces basées sur cette approche offrent souvent un moyen visuel (icônes, schéma, dessin) afin d'exprimer l'information recherchée.

Une première famille de cette approche offre une relative liberté à l'utilisateur pour dessiner ses configurations spatiales. Spatial-Query-By-Sketch [5] [6] et le langage Sketch! [14] sont des exemples type de cette approche.

La deuxième famille de cette approche représentée par Cigales [3] et LVIS [1] [2], offre un ensemble d'icônes schématisant des objets géographiques et des relations spatiales (inclusion, intersection, adjacence,...) qu'on peut combiner pour construire la requête.

A l'issue des travaux précités et en se basant sur la métaphore des champs électromagnétiques l'utilisation des langages visuels a été étendue aux données de type champ continu scalaire à deux dimensions avec les travaux de Robert Laurini et Luca Paolino [12]. Dans ces recherches la dimension temporelle n'a pas été prise en compte.

Basés sur le fait que les informations géographiques sont des informations visuelles par nature et qu'en réalité les phénomènes naturels évoluent continuellement dans le temps et dans l'espace, notre travail se fixe comme objectif la définition d'un langage visuel pour l'interrogation des bases de données à champs continus.

Le recours aux langages visuels pour interroger les données de type champ continu est dû à leur aspect conviviale et simple d'utilisation. Leur mode d'interrogation est simple et bien adapté aux utilisateurs occasionnels non experts.

Les paragraphes qui suivent du présent article décrivent notre approche pour aborder cette problématique.

3 Approche visuelle préconisée

Un utilisateur qui veut interroger une base de données champs continus a déjà délimité son domaine d'action, autrement dit, il a déjà réalisé sa requête en termes provenant du processus de son domaine d'expertise.

Le travail consiste à offrir à cet utilisateur la possibilité de traduire cette requête en termes mis à sa disposition. On appellera l'effort de traduction : la formulation de la requête, et l'ensemble des termes mis à sa disposition : la représentation de la réalité du centre d'intérêt de l'utilisateur.

3.1 Spécifications

L'étude portée sur le problème de la formulation de la requête nous a permis de classer les requêtes en deux catégories : les requêtes de manipulation (Fig.1) et les requêtes de sélection (Fig.2).

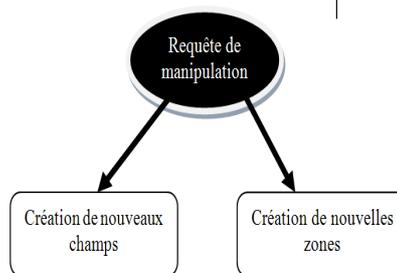


Fig. 1. Requête de manipulation

Pour ce second type de requêtes nous avons considéré :

- les requêtes unaires à un champ scalaire ou vectoriel. Par exemple « quelle est la moyenne de température sur le Rhône ? »
- et les requêtes binaires portant sur deux champs : soient scalaires, soient vectoriels soient mixtes. Un exemple typique de requête binaire à champ scalaire est « quelle est la quantité d'eau tombée sur un rayon de 100 km autour de Tlemcen sachant que l'on a un champ de pluie par pays ? »

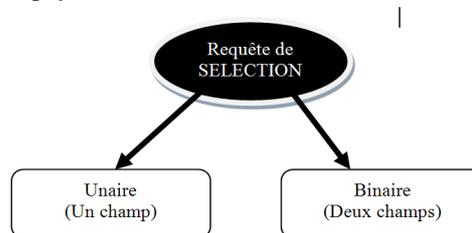


Fig. 2. Requête de manipulation

Dans notre environnement visuel, la formulation de la requête que celle-ci soit une requête de sélection ou de manipulation est réalisée à base de sélections successives d'icônes et de métaphores. Au cours de cette formulation la représentation visuelle du centre d'intérêt de l'utilisateur est construite progressivement.

Les icônes représentent les objets de la base de données d'une part et les métaphores représentent les types d'objets et les relations nécessaires à l'expression des contraintes spatiales et temporelles dans la requête d'autre part.

D'un autre côté nous avons défini pour chaque objet de la base de données un identifiant d'objet, un identifiant thématique (en général constitué du type thématique de l'objet et de son nom, un localisant spatial, un localisant temporel et un ensemble d'attributs thématiques (essentiellement, à valeurs alphanumériques)).

Nous supposons ici que tous les objets possèdent un localisant spatial déterminant leur type spatial et leur géométrie (sous la forme d'une séquence de points de l'espace).

Nous considérons dans cette étude les objets ponctuels, linéaires et surfaciques. Les objets volumiques ne sont pas pris en considération par simplification, bien qu'ils puissent tout à fait être intégrés à notre langage comme une extension naturelle.

La représentation visuelle de ces types spatiaux (linéaire et surfacique) est donnée tel que les objets ponctuels sont assimilés à des objets surfaciques de surface nulle et donc leur représentation visuelle est la même que pour celle des objets surfaciques.

Enfin, notons que certains objets n'ont pas de localisation spatiale significative. Les objets mobiles, par exemple, sont en perpétuel mouvement. Ces objets sont alors assimilés à des objets ponctuels et leur localisation spatiale est liée à leur localisation temporelle.

Nous supposons aussi que tous les objets possèdent un localisant temporel exprimé sous la forme d'un intervalle de temps délimité par deux instants selon un axe temporel donné. Il n'y a qu'un seul type temporel qui est l'intervalle. Un instant est considéré comme étant un intervalle de durée nulle.

Un type d'objet de la base de données sera représenté visuellement par son type thématique et son type spatial. De même un objet est représenté par son identifiant thématique et son type spatial. Le type thématique est visualisé sous la forme d'une icône. L'identifiant thématique est constitué de l'icône du type thématique et d'une valeur textuelle qui labellise cette icône.

La représentation visuelle du type spatial est alors accolée à celle du type thématique (resp. de l'identifiant thématique) pour donner la représentation visuelle complète du type d'objet (resp. de l'objet).

Le type thématique est le nom logique d'une table ou d'une couche de carte de la base de données. L'identifiant thématique désigne un objet appartenant à la table ou la couche de carte dont le nom correspond à la valeur de son type thématique. Ainsi, dans notre base de données exemple, les types thématiques, et donc les tables, sont Forêts, Lacs, Ville, routes, etc.

Une requête met en jeu des objets ou types d'objets avec un ou plusieurs critères. Les objets ou types d'objets de la requête constituent ses opérands ; les critères, des opérateurs s'appliquant sur ses opérands.

Les requêtes formulées sont qualifiées de thématiques, spatiales, temporelles ou spatio-temporelles selon le contenu de leur critère de sélection. Une requête thématique (resp. spatiale, temporelle, spatio-temporelle) portent sur les valeurs des

attributs thématiques (resp. localisant spatial et localisant temporel) des objets mis en jeu dans cette requête.

Les spécifications précitées de notre système nous ont permis de réaliser notre environnement visuel qui comporte un éditeur de requêtes visuelles au centre duquel une zone de travail est réservée comme espace de construction d'une requête à partir des éléments visuels sélectionnés étape par étape (Fig. 3).

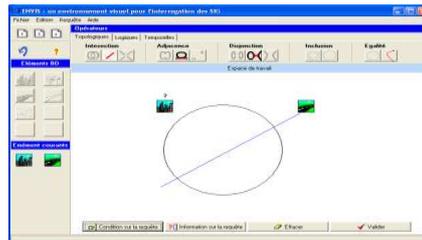


Fig. 3. Exemple de formulation visuelle d'une requête géographique

A la fin de la formulation de la requête celle-ci peut être enregistrée pour une utilisation ultérieure, effacée en cas d'erreur ou validée pour être traduite dans le langage cible SQL. D'autres conditions attributaires peuvent enrichir la requête visuelle par la saisie d'un formulaire.

5 Conclusion

L'approche figurative préconisée dans notre étude permet de remédier aux problèmes de récupération de données géographiques organisées de manière sensée dans les bases de données champs continus.

En effet, une requête visuelle permet de raccourcir la distance entre le modèle mental de l'utilisateur et la représentation de la réalité proposée par la machine. Elle réduit la dépendance de l'utilisateur avec le langage naturel dans sa forme textuelle et la difficulté des langages de commandes comme SQL.

Elle augmente le taux de performances par une interface plus naturelle et plus intuitive qui permet aux utilisateurs de définir de nouveaux objectifs en mettant à leur disposition des moyens et des opérations nécessaires pour leur permettre de récupérer de manière fidèle un ensemble de données. Ce mode d'interrogation offre un moyen simple et facile à l'utilisateur final qui exprime sa requête à l'aide d'icônes et de métaphores.

La représentation visuelle est une représentation adaptée à l'utilisateur des bases de données champs continus et un axe très prometteur pour remédier à la problématique de l'interrogation de ces dernières.

References

1. C. Bonhomme, M.A. Aufaure: "A Visual Language for High Spatial Data Management", *Visual'99, Third International Conference on Visual Information Systems*, Amsterdam, pp. 325-332, Springer Verlag, Lecture Notes in Computer Science 1614
2. C. Bonhomme, MA Aufaure, R. Laurini, C. Trépied: "A Visual Language for Querying Spatio-Temporal Databases". Proc. of the *7th International Symposium on GIS, ACMGIS'99*, Kansas City, November 5-6, 1999, edited by C. Bauzer-Medeiros, ACM-Press, pp. 34-39.
3. D. Calcinelli and M. Mainguenaud "Cigales: a visual query language for a geographical information system: the user interface". *Journal of Visual Languages and Computing*, (5):113-132, 1994.
4. M.J Egenhofer: "Extending SQL for Cartographic Display" *Cartography and Geographic Information System*, pp. 230-245, 1991
5. M. Egenhofer: "Spatial-query-by-sketch". In M. Burnett and W. Citrin eds., *VL'96 : IEEE Symposium on Visual Languages*, Boulder, CO, p. 60-67, September 1996.
6. M. Egenhofer. "Query processing in spatial-query-by-sketch". *Journal of Visual Languages and Computing*, (8) :403-424, 1997.
7. M. Gaio, "Traitements de l'information Géographique : représentations et structures", Université de Caen-Basse-Normandie, Habilitation à diriger des recherches, 2001
8. M. Gaio, J. Madelaine, C. Turbout, "De la gestion à la réutilisation d'information géographique hétérogène multi-sources sur Internet", Actes des *journées Cassini 2002*, France, pp. 1-20
9. S. Gordillo: "Modélisation et manipulation de phénomènes continus spatio-temporels". Thèse de Doctorat Informatique, 12 octobre 2001. Université Claude Bernard Lyon I.
10. A. Lbath, "AIGLE: un environnement visuel pour la conception et la génération automatique d'applications géomatiques", Thèse de Doctorat en Informatique, INSA Lyon, 1997.
11. R. Laurini, D. Pariente: "Towards a Field-oriented Language: First Specifications". In "Geographic Objects with Indeterminate Boundaries" Edited by *Burrough and Frank, Taylor and Francis*, 1996. pp 225-236
12. R. Laurini, L. Paolino, M. Sebillio, G. Tortora, G. Vitiello "Phenomena - A Visual Query Language for Continuous Fields", *Proceedings of the 11th ACM-GIS'03 Conference*, New Orleans, Louisiana, November 7-8, 2003, pp. 147-153.
13. R. Laurini, L. Paolino, M. Sebillio, G. Tortora, G. Vitiello: "Dealing with geographic continuous fields: the way to a visual GIS environment". In *Proceedings of AVI 2004: Conference on Advanced visual interfaces, AVI 2004*, Gallipoli, Italy, May 25-28, 2004, pp. 336-343.
14. B. Meyer: "Towards new metaphors for visual query languages for spatial information systems". In *Proc. of the first International Workshop on Interfaces to Database Systems*, pages 113-135. R. Cooper, Springer Verlag, Glasgow, UK 1993.
15. C. Mostaccio: "Organisation Physique des bases de données pour les champs continus". Thèse de Doctorat Informatique, 21 novembre 2003. Université Claude Bernard Lyon I.
16. D. Pariente : "Estimation, Modélisation et Langage de Déclaration et de Manipulation de Champs Spatiaux Continus", Thèse de doctorat Ingénierie Informatique, 1994. INSA de Lyon.
17. S. Ploux, H. Ji: "A model for matching semantic maps between languages (French/English, English/French)", *Computational Linguistics*, 29 (2), 155-178, 2003

18. N. Roussopoulos, C. Faloutsos, T. Sellis: "An efficient pictorial database system for PSQL" *IEEE Transactions on Software Engineering*, 14 (5), 639-650, 1988
19. R.T. Snodgrass, M.H. Bhlen, C.S. Jensen, A. Steiner: "Transitioning temporal support in TSQL2 to SQL3". In: *Temporal database management* (ed. Jensen, C.S.), pp.615-663.1998
20. M. Thériault, « Système d'Informations Géographiques: Concepts Fondamentaux », Notes de cours N° 12, LATIG, Département de Géographie, Université Laval, 1996
21. C.D. Tomlin. "Digital cartographic modelling techniques in environmental planning". Ph.D., School of Forestry and Environmental Studies, New Haven, CT, Yale University.1983
22. C.D. Tomlin. "Geographic information systems and cartographic modelling". Englewood Cliffs, NJ, Prentice Hall. 249 p. 1990. »
23. C. Turbout, "Construction d'hypertexte et recherche d'informations hétérogènes : la spécificité de l'information géographique", Thèse de Doctorat en Informatique, Université de Caen, France, 2002.