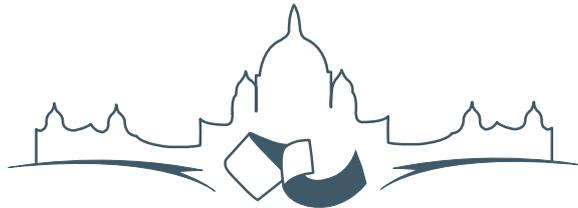

Journal de l'OSGeo

Le journal de la Fondation Open Source Geospatial

Volume 3 / Décembre 2007



2007 FREE AND OPEN SOURCE SOFTWARE
FOR GEOSPATIAL (FOS4G) CONFERENCE
VICTORIA CANADA  SEPTEMBER 24 TO 27, 2007

Compte rendu du FOS4G 2007

Intégration & Développement

- Portable GIS : SIG sur une clé USB
- Génération Automatique d'Applications SIG / Base de données sur Internet
- db4o2D - Extension de Base de données Orientée Objet pour les types géospatiaux 2D
- Google Summer of Code pour la géomatique

Intérêt thématique

- Approche générique pour la gestion de standards de métadonnées
- Vers des services web dédiés à la cartographie thématique
- Interopérabilité pour les données géospatiales 3D
- Un Service Web orienté modèle pour une interopérabilité sémantique améliorée
- Spatial-Yap : un système de base de données spatialement déductif

Études de cas

- Le Projet DIVERT : Développement de Télématiques Inter-Véhicules Fiabiles
 - GRASS et la Modélisation des Risques Naturels
 - Une Base de Données Spatiales pour l'Intégration des Données du Projet de Gestion des Ressources Naturelles du Rondonia
 - GeoSIPAM : Logiciel libre et Open Source appliqué à la protection de l'Amazonie brésilienne
 - Le Système de Suivi de la Déforestation Amazonienne
-

Spatial-Yap : un système de base de données spatiales déductive

David Vaz et Michel Ferreira, traduit par Sidone Christophe

Introduction

Le paradigme des bases de données déductives est une extension de celui des bases de données traditionnelles avec des possibilités de déduction. La connaissance n'est pas seulement représentée par extension¹⁹, mais également par des règles logiques intentionnelles. Une approche classique dans la construction d'un tel système de base de données déductif est de coupler un système de programmation logique avec un système de base de données relationnelle.

Le langage de requête Datalog (1), réduit les données attributs à un champ de valeurs atomiques, tel que nombres et chaînes de caractères, lesquels sont les données typiques stockées dans une base de données. Cependant, les bases de données actuelles gèrent bien plus de structures de données, telles que les attributs géométriques des relations spatiales.

Dans cet article, nous décrivons l'extension du compilateur Yap Prolog (2) un système de programmation gratuit, open-source, pour qu'il gère la donnée spatiale, fournissant une solution dernier cri pour la modéliser, la requêter et l'extraire. Une proposition pour étendre Datalog vers Spatial Datalog a été décrite dans la littérature, dans le cadre des bases de données contraintes (3).

L'approche suivie dans Spatial Yap est différente et plus proche de la communauté des bases de données spatiales, car elle est basée sur des notions spatiales, plutôt que des inégalités polynomiales. Spatial-Yap peut construire des relations spatiales logiques entre données vectorielles et procure un environnement de programmation déclaratif pour sa gestion, intégrant, par exemple, des spécifications de récursivité, inhérente aux relations topologiques, et par un puissant ADT, comme outil logique. Bien que Spatial-Yap est bien plus orienté sur le déclaratif que sur l'efficacité, le système est capable d'explorer les fonctions avancées de Yap, comme un traducteur Prolog vers SQL et un moteur basé sur des essais, pour en améliorer les performances.

Programmation Logique et Programmation Logique Inductive

La Programmation Logique (LP), dont Prolog est le langage canonique, tente d'implémenter l'idée de Colmerauer et Kowalski selon laquelle le calcul est contrôlé par inférence (4). La motivation du paradigme de la LP est de séparer *ce que* doit faire un programme du *comment* il doit le faire. Cela a été résumé par la devise de Kowalski :

algorithme = logique + contrôle.

Les programmes Prolog utilisent la logique pour exprimer un problème et s'appuient sur le système Prolog pour exécuter cette spécification. Prolog implémente un sous-ensemble logique de premier ordre comme les clauses de Horn.

$A :- B_1, B_2, \dots, B_n.$

ce qui veut dire : A est Vrai si B1 est Vrai et B2 est Vrai ... et Bn est Vrai. Ces règles donnent une interprétation procédurale qui se lit ainsi :

pour résoudre(exécuter) A résoudre(exécuter) B1 et résoudre(exécuter) B2 ... et résoudre(exécuter) Bn.

L'interprétation procédurale précise utilisée dans les programmes Prolog est une forme réduite de la résolution SLD (5).

Les systèmes de bases de données déductifs sont des systèmes de gestion de bases de données qui sont aussi conçus autour d'un modèle logique de données et dont le langage de requêtes est une version *set-oriented* de Prolog, connu sous le nom de Datalog. Les relations sont naturellement pensées comme la *valeur* du prédicat logique et la puissance expressive des expressions logiques est utilisée pour requêter de telles relations. La part déductive de tels systèmes vient du fait que les moteurs de Programmation Logique utilisent des *intentions* (ou compréhensions), pour exprimer des propriétés, et sont capables de matérialiser ces intentions dans des connaissances *extensionnelles* (tuples relationnels ou faits).

Ce processus de déduction, qui va d'intentions en extensions, est d'un point de vue calculatoire plus simple que le processus inverse qui consiste à aller d'extensions en intentions. Cependant, être capable de dériver une représentation intentionnel depuis des données extensionnelles, en inférant une

¹⁹Définition des termes : http://en.wikipedia.org/wiki/Extensional_definition

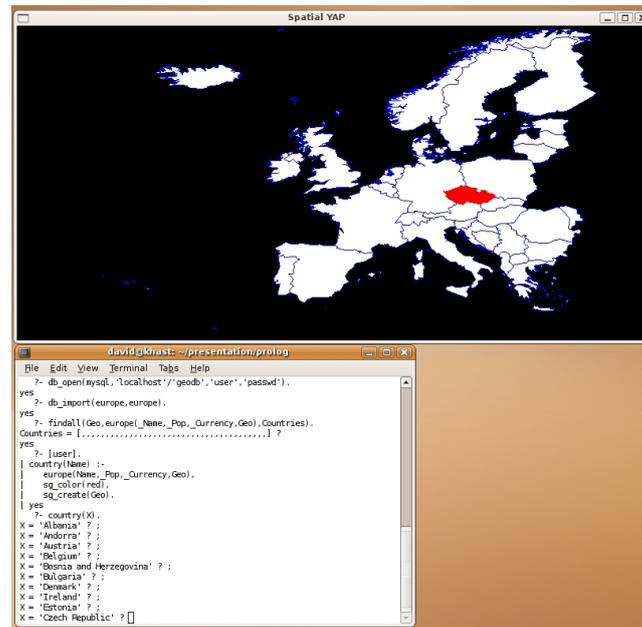


FIG. 1: Exemple de Spatial-Yap.

règle générale à partir d'exemples, est aussi crucial. C'est le but du paradigme de la Programmation Logique en tant que Programmation Logique Inductive (PLI) (6).

La Programmation Logique Inductive (PLI) est un domaine de recherche à l'intersection de l'Apprentissage Automatique et de la Programmation Logique. Les systèmes de PLI développent des prédicats à partir d'exemples et d'une base de connaissances, dont on déduit un programme logique qui confirme les exemples positifs et infirme les exemples négatifs. Pour déduire une théorie avec les propriétés attendues, beaucoup de systèmes de PLI suivent une sorte d'approche *generation-et-test* pour balayer *l'ensemble des hypothèses*. Une caractéristique importante de l'approche de la PLI pour la fouille de données est d'être multi-relationnelle, capable de formuler des théories qui impliquent des données ayant plusieurs relations, alors que beaucoup de techniques importantes de fouilles de données sont seulement capables de dégager des structures pour une unique relation. Ceci est particulièrement intéressant pour la fouille de données spatiales, qui est multi-relationnel (ou multi-couches) par définition.

Un système de base de données spatiales déductive

Spatial Yap résulte d'une interface complexe entre plusieurs composants. Les deux composants principaux sont le système Yap Prolog et le SGBD MySQL, qui sont couplés par une interface MYDDAS (7) (pour Mysql/Yap Deductive Database System). Cette interface est responsable du couplage des deux systèmes comme le montre la Figure 2. MYDDAS traduit les requêtes logiques en requêtes SQL de façon transparente, implémente la conversion des attributs MySQL en notions Yap et explore le moteur YapTab pour résoudre des requêtes récursives impliquant la base de données. Le niveau de sophistication de cette interface est très élevé : la recherche de tuples relationnels est directement implémentée dans le choix des points WAM, supportant les opérateurs simplifiés (8).

Afin de construire la base de données spatiales déductive, nous avons étendu l'interface MYDDAS afin qu'elle intègre des types géométriques MySQL. Deux composants supplémentaires sont fondamentaux pour construire Spatial Yap : une librairie d'opérateurs spatiaux basée sur la librairie classique GEOS et un composant de visualisation.

Quand on manipule des données spatiales il est essentiel d'être capable de représenter ces données graphiquement. Représenter une carte comme un en-

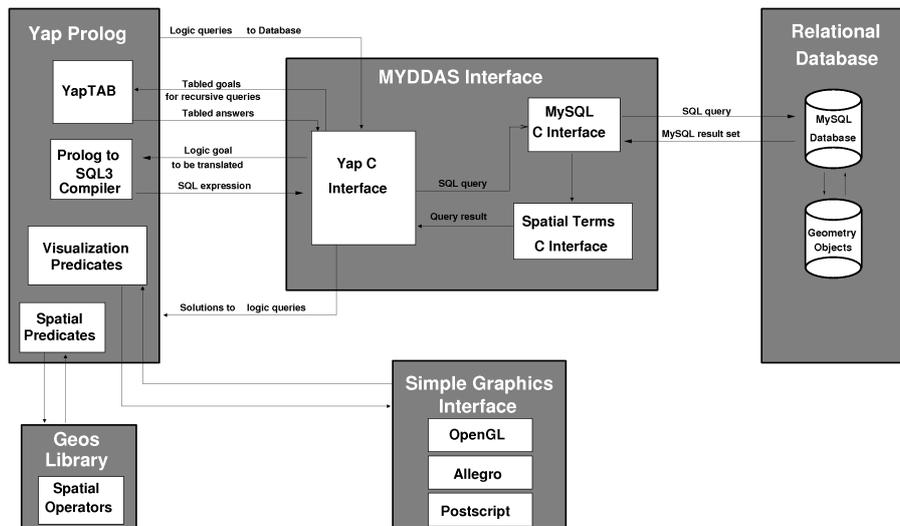


FIG. 2: Modèle Spatial Yap.

semble de notions Prolog est visuellement inacceptable du point de vue d'un utilisateur de SIG. Plus importante encore est la représentation d'une opération spatiale, comme l'intersection de deux polygones d'un point de vue graphique. Les analyses spatiales réalisées par l'utilisateur et la représentation des ensembles de résultats des requêtes spatiales nécessitent l'ajout d'un composant de visualisation à tout système de base de données spatiales. Nous aurions pu utiliser un des FOSS existants, comme MapServer, mais nous avons besoin de quelque chose de plus simple qui fonctionne à un haut-niveau graphique, étroitement lié avec le haut-niveau textuel de notre système Prolog. L'interaction entre ces deux fonctionnements haut-niveau était notre objectif principal, plutôt qu'un affichage graphique sophistiqué. La Figure ?? montre une copie d'écran de l'interaction entre les deux parties haut-niveaux de Spatial-Yap.

Applications en cours

Dans cette section nous décrivons deux projets utilisant Spatial-Yap en cours.

Étude du comportement du trafic dans la ville de Porto

Le but de ce projet est d'étudier le comportement du trafic dans la ville de Porto, deuxième plus importante du Portugal, avec un réseau routier de 965 km, présenté Fig. 3. Nous nous intéres-

sons à la compréhension de facteurs affectant le trafic, pas uniquement concernant le temps et le jour, mais aussi incluant des entités géographiques intrinsèques, comme la présence d'écoles sur un tronçon de route et leur influence sur la congestion du trafic suivant certaines plages horaires. Un des objectifs plus ambitieux est la dérivation de la couche de signalisation routière, incluant les lumières et la localisation des stops et basée sur des modèles de mobilité mais aussi l'inférence des possibles destinations des conducteurs pouvant automatiquement activer les systèmes de navigation basés sur un fond des routes usuels.



FIG. 3: Road network of Porto.

Des algorithmes de routage et l'affichage des itinéraires calculés sont aussi implémentés en utilisant Spatial-Yap et son moteur. Les systèmes de PLI fournissent le support d'inférence sur les données géospatiales, comme les logs GPS.

Correction de la Classification Automatique des Forêts basée sur l'Analyse Spatiale

Un autre projet intéressant où nous utilisons Spatial-Yap vise à la surveillance globale des changements de biodiversité (9). Les gouvernements ont comme objectif ambitieux de réduire la baisse de la biodiversité d'ici 2010, et les scientifiques sont confrontés maintenant au défi de l'accès aux progrès réalisés dans ce sens. Le projet européen Corine Land Cover²⁰ (CLC) fournit des données pour deux années différentes (1990 et 2000), utilisant 44 classes d'occupation du sol. Ces données sont acquises à partir d'images satellites, grâce à la vectorisation de chacune des 44 classes d'occupation du sol en polygones, réalisée automatiquement, par reconnaissance de couleurs. Malheureusement les classes CLC ne sont pas les plus appropriées pour surveiller la biodiversité. Par exemple, la base CLC courante n'a que 3 classes pour la forêt (feuillus, conifères et mixtes). Par conséquent, une augmentation observée dans la zone forestière de feuillus pourrait être due à une augmentation de la superficie d'une plantation d'espèces exotiques, telles que l'Eucalyptus globulus, ou à une augmentation du nombre d'arbres dans les forêts natives de feuillus, les deux phénomènes ayant des implications différentes en terme de biodiversité. Le groupe de biologistes avec qui nous travaillons possède des cartes régionales de la région de l'Alto Minho au Portugal (see Fig. 4), aussi couverte par CLC. Ces cartes régionales sont réalisées sur place avec des techniques de cartographie lentes et coûteuses et l'identification sur place des espèces en forêt permet un meilleur niveau de détail dans la liste des classes. Notre projet essaie d'utiliser ces cartes régionales afin de dériver un ensemble de règles spatiales logiques qui permettent la caractérisation détaillée des données CLC pour la surveillance de la biodiversité.

²⁰Le Projet Européen Corine Land Cover (CLC) : <http://terrestrial.eionet.eu.int/CLC2000>

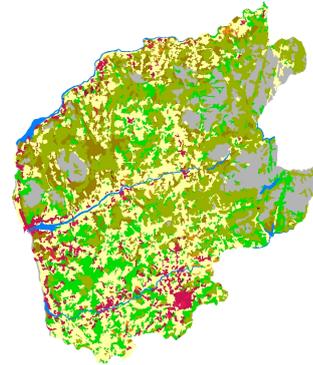


FIG. 4: Alto Minho Map.

Perspectives

La migration de MySQL à PostgreSQL est en cours. Notre but est d'étendre Prolog comme compilateur SQL, afin d'être capable de transférer dans le système de base de données la convergence des buts logiques, ce qui peut être préférable à l'indexation spatiale qui n'est pas actuellement disponible dans Spatial-Yap.

Spatial-Yap peut être téléchargé depuis myddas.dcc.fc.up.pt. Un manuel utilisateur et plusieurs articles avec des présentations plus approfondies de Spatial-Yap sont aussi disponibles sur cette même page.

Remerciements

Ce travail a été partiellement soutenu par MYDDAS (POSC/EIA/59154/2004) et par des fonds accordés à LIACC à travers le Programa de Financiamento Plurianual, Fundação para a Ciência e Tecnologia and Programa POSC. David Vaz est subventionné par une bourse de doctorat de type FCT PhD

grant SFRH/BD/29648/2006.

Bibliographie

- [1] J. D. Ullman. *Principles of Database and Knowledge-Base Systems*. Computer Science Press, 1989.
- [2] V. Santos Costa, L. Damas, R. Reis, and R. Azevedo. *YAP User's Manual*. Available from <http://www.ncc.up.pt/~vsc/Yap+>.
- [3] P. C. Kanellakis, G. M. Kuper, and P. Z. Revesz. Constraint query languages. *J. Comput. Syst. Sci.*, 51(1) :26–52, 1995.
- [4] R. Kowalski. Predicate Logic as a Programming Language. In *Information Processing*, pages 569–574. North-Holland, 1974.
- [5] J. W. Lloyd. *Foundations of Logic Programming*. Springer-Verlag, 1987.
- [6] S. Muggleton. Inductive Logic Programming. In *Conference on Algorithmic Learning Theory*, pages 43–62. Ohmsma, 1990.
- [7] T. Soares, M. Ferreira, and R. Rocha. The MYDDAS Programmer's Manual. Technical Report DCC-2005-10, Department of Computer Science, University of Porto, 2005.
- [8] T. Soares, R. Rocha, and M. Ferreira. Generic Cut Actions for External Prolog Predicates. In P. V. Hentenryck, editor, *Proceedings of the 8th International Symposium on Practical Aspects of Declarative Languages, PADL'2006*, number 3819 in LNCS, pages 16–30, Charleston, South Carolina, USA, January 2006. Springer-Verlag.
- [9] H. M. Pereira and H. D. Cooper. Towards the global monitoring of biodiversity change. *Trends in Ecology & Evolution*, 21(3) :123–129, March 2006.

David Vaz
DCC-FC & LIACC, University of Porto
[davidvaz AT dcc.fc.up.pt](mailto:davidvaz@dcc.fc.up.pt)

Michel Ferreira
DCC-FC & LIACC, University of Porto
[michel AT dcc.fc.up.pt](mailto:michel@dcc.fc.up.pt)

The [Open Source Geospatial Foundation](#), or OSGeo, is a not-for-profit organization whose mission is to support and promote the collaborative development of open geospatial technologies and data. The foundation provides financial, organizational and legal support to the broader open source geospatial community. It also serves as an independent legal entity to which community members can contribute code, funding and other resources, secure in the knowledge that their contributions will be maintained for public benefit. OSGeo also serves as an outreach and advocacy organization for the open source geospatial community, and provides a common forum and shared infrastructure for improving cross-project collaboration.

Publié par l'OSGeo, le Journal de l'OSGeo a pour objectif de publier les résumés des conférences, étude de cas et introduction, et les concepts liés à l'open source et aux logiciels en géomatique.

Équipe éditorial :

- Angus Carr
- Mark Leslie
- Scott Mitchell
- Venkatesh Raghavan
- Micha Silver
- Martin Wegmann

Rédacteur en Chef :

Tyler Mitchell - [tmitchell AT osgeo.org](mailto:tmitchell@osgeo.org)

Remerciements

Divers relecteurs & le project GRASS News

Le *Journal de l'OSGeo* est une publication de la *Fondation OSGeo*. La base de ce journal, le source des styles L^AT_EX 2_ε a été généreusement fournie par le bureau d'édition de GRASS et R.



Ce travail est sous licence Creative Commons Attribution-No Derivative Works 3.0 License. Pour lire une copie de cette licence, visitez : creativecommons.org.



All articles are copyrighted by the respective authors — contact authors directly to request permission to re-use their material. See the OSGeo Journal URL, below, for more information about submitting new articles.

Journal en line : <http://www.osgeo.org/journal>

OSGeo Homepage : <http://www.osgeo.org>

Courrier postal : OSGeo

PO Box 4844, Williams Lake,

British Columbia, Canada, V2G 2V8

Téléphone : +1-250-277-1621



ISSN 1994-1897

This PDF article file is a sub-set from the larger
OSGeo Journal. For a complete set of articles
please the Journal web-site at:

<http://osgeo.org/journal>