GRASS/ OSGeo-News

Open Source GIS and Remote Sensing informations Volume 1, June 2004

Spatial-Yap: A Spatio-Deductive Database System

by David Vaz and Michel Ferreira

Introduction

The paradigm of Deductive Databases extends traditional databases with deduction abilities. Knowledge in not only represented extensionally but also through intensional logic rules. A typical approach in building such deductive database systems is to couple a logic programming system with a relational database system.

The original query language, Datalog (1), restricted attribute data to ground atomic values, such as numbers and strings, which were the typical data stored in databases. However, current databases store much more structured data, such as the geometric attributes of spatial relations.

In this paper we describe the extension of the Yap Prolog (2) compiler, a free, open-source logic programming system, in order to handle spatial data, providing a state-of-the-art solution for its modeling, querying and mining. A proposal of extending Datalog to Spatial Datalog has been described in the literature, in the framework of Constraint Databases (3). The approach followed in Spatial-Yap is different and closer to the spatial databases community, as it is based on spatial terms, rather than on polynomial inequalities. Spatial-Yap can build spatial logic terms from vectorial data in spatial relations, and provides a highly declarative programming environment for its handling, supported, for instance, by the natural specification of recursion, inherent to topological relationships, and by a powerful ADT, as is the logic term. Although the current focus in Spatial-Yap is much more on declarativity than on efficiency, the system is able to explore advanced features of Yap, such as a Prolog to SQL translator and a tabling engine based on tries, to improve performance.

Contents of this volume:

Logic Programming and Inductive Logic Programming

Logic programming (LP), of which Prolog is the canonical language, is an attempt to implement Colmerauer and Kowalski's idea that computation is controlled inference (4). The motivation for the LP paradigm is to separate the specification of *what* the program should do from *how* it should be done. This was summarized by Kowalski's motto:

$$algorithm = logic + control.$$

Prolog programs use the logic to express the problem and rely on the Prolog system to execute this specification. Prolog implements a subset of first order logic known as Horn clause logic. A Prolog program is a set of relational rules of the form:

`A :- B1, B2, ..., Bn.`

meaning: `A` is `True` if `B1` is `True` and `B2` is `True` ... and `Bn` is `True`. These rules are given a procedural interpretation which reads as:

to solve(execute) `A` solve(execute) `B1` and solve(execute) `B2` ... and solve(execute) `Bn`.

The precise procedural interpretation used in the execution of Prolog programs is a restricted form of SLD-resolution (5).

Deductive database systems are database management systems which are also designed around a logic model of data and whose query language is a set-oriented version of Prolog, known as Datalog. Database relations are naturally thought of as the *value* of a logical predicate and the high expressive power of logic expressions is used to query such relations. The deductive part of such systems comes from the fact that the logic programming engines take *intentions* (or comprehensions), which express properties, and are able to materialize these intentions in *extensional* knowledge (relational tuples or facts).

This process of deduction, which goes from intentions to extensions, is computationally much simpler than the reverse process of going from extensions to intentions. However, being able to derive an intentional representation from extensional data, inferring a general rule from examples, is also crucial. This is the goal of a logic programming paradigm known as Inductive Logic Programming (ILP) (6).

ILP is a research area formed at the intersection of Machine Learning and LP. ILP systems develop predicate descriptions from examples and background knowledge, thus deriving an hypothesized logic program which entails all the positive and none of the negative examples. To derive a theory with the desire properties, many ILP systems follow some kind of *generate-and-test* approach to traverse the *hypotheses space*. An important characteristic of the ILP approach to data mining is that it is multi-relational, being able to formulate theories which involve data in several relations, while many important data mining techniques are only able to look for patterns in a single relation. This is particularly useful for spatial data mining, which is inherently multi-relational (or multi-layered).

A Spatio-Deductive Database System

Spatial Yap results from a sophisticated interface between several components. The two main components are the Yap Prolog system and MySQL RDBMS, which are coupled through the MYDDAS interface (7) (Mysql/Yap Deductive DAtabase System). This interface is responsible for coupling these two systems, as illustrated in Figure 1. MYDDAS transparently translates logic queries into SQL statements, implements the conversion into Yap terms of MySQL attributes and explores the YapTab tabling engine for solving recursive queries involving database goals. The level of sophistication of this interface is very high, with the fetching of relational tuples being implemented directly in WAM choicepoints, supporting pruning operators (8).

To build the spatial deductive database system we extended the MYDDAS interface to support MySQL Geometry Types. Two more components are fundamental to build Spatial Yap: a spatial operators library based on the well known Geos library and a visualization component.

When dealing with spatial data it is essential to be able to graphically represent such data. Representing a map as a set of Prolog terms is visually unacceptable, from the point of view of a GIS user. Even more important is the representation of a spatial operation, such as the intersection of two polygons, in a graphical way. User-driven spatial analysis and the representation of spatial queries result sets require a visualization component to be added to any spatial database system. Here we could have used one of the existing FOSS, such as MapServer, but we rather needed something simpler that worked as a graphical spatial top-level, tightly coupled to the text top-level of our Prolog system. Interaction between these top-levels was our main goal, rather than sophisticated graphical display. Figure 2 show a screenshot of the interaction between the two top-levels of Spatial-Yap.

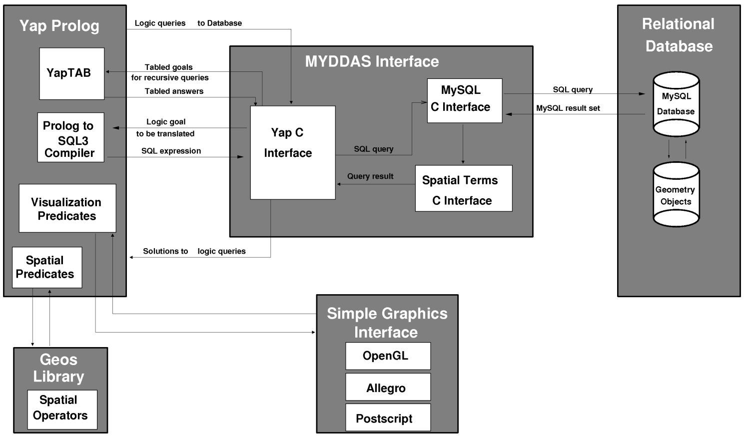

Figure 1: Spatial Yap Blueprint.

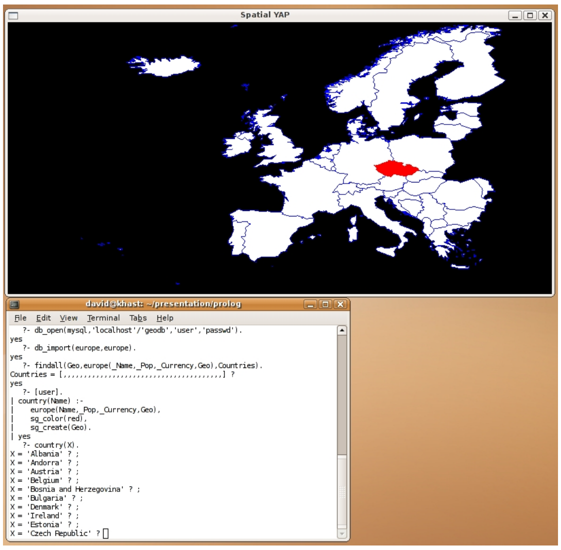

Figure 2: Spatial-Yap example.

Current Applications

In this section we describe two undergoing projects where we are applying Spatial-Yap.

Study of Traffic Behaviour in the City of Porto

The aim of this project is to study traffic behavior in the city of Porto, second largest in Portugal, with a road network totaling 965 kms, shown in Fig. 3. We are interested in understanding factors affecting traffic, not only time and day related, but also including intrinsic geographic entities, such as the presence of a school in a street segment and its influence in traffic congestion time-slots. More ambitious goals include the automatic derivation of a road signalization layer, including traffic lights and stop signs locations, based on mobility patterns, or the inference of likely destinations of drivers that can automatically activate navigation systems, based on a background of usual routes.

Figure 3: Road network of Porto.

Routing algorithms and displaying of computed routes are also implemented using Spatial-Yap and its tabling engine. ILP systems provide the support for inference over geospatial data, such as GPS logs.

Correction of Automatic Classification of Forests based on Spatial Analysis

Another interesting project where we are using Spatial-Yap aims at the global monitoring of biodiversity change (9). Governments have set the ambitious target of reducing biodiversity loss by the year 2010, and scientists now face the challenge of accessing the progress made towards this target. The European Corine Land Cover (CLC) project (`http://terrestrial.eionet.eu.int/CLC2000`) provides data for two different years (1990 and 2000), using 44 land-cover classes. This data is obtained from satellite images and the vectorization in the polygons of each of the 44 land-cover classes is done automatically, based on color recognition. Unfortunately, the land-cover classes of CLC are not the most appropriate to monitor biodiversity. For instance, currently CLC has only three classes for forest (broad-leaved, coniferous and mixed); therefore, an observed increase in broadleaved forest area could be due to an increase in plantation area of an exotic species, such as Eucalyptus globulus, or an increase in native broad-leaved forest, two phenomena with different implications for biodiversity. The group of biologists with whom we are working has detailed regional maps from the area of Alto Minho, in Portugal (see Fig. 4), also covered by CLC. These regional maps are done based on expensive and slow on-site mapping techniques and on-site identification of forest species, allowing a much higher detail on the list of classes. Our project is trying to use these detailed regional maps to derive a set of spatial logic rules that allow the detailed characterization of CLC data for biodiversity monitoring. We are using Spatial-Yap and the APRIL ILP system over data-sets created based on the intersection of the regional maps and CLC maps. The inducted rules can then be used to improve the categorization of new CLC data, allowing its use for biodiversity monitoring.

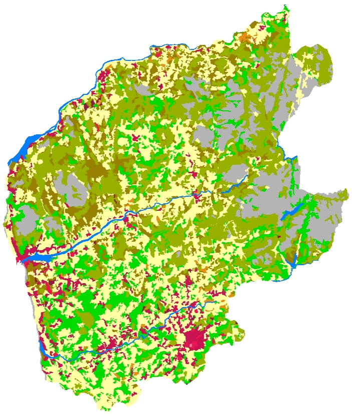

Figure 4: Alto Minho Map.

Future Work

A migration from MySQL to PostgreSQL is undergoing. Our goal is to extend the Prolog to SQL compiler in order to be able to transfer to the database system conjunctions of logic goals, that can take advantage of spatial indexing that currently is not available in Spatial-Yap.

Spatial-Yap can be downloaded from `http://myddas.dcc.fc.up.pt`. A users manual and several papers with deeper presentations of Spatial-Yap are also available from the same webpage.

Acknowledgments

This work has been partially supported by MYDDAS (POSC/EIA/59154/2004) and by funds granted to LIACC through the Programa de Financiamento Plurianual, Fundação para a Ciência e Tecnologia and Programa POSC. David Vaz is funded by FCT PhD grant SFRH/BD/29648/2006.

Bibliography

[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`

Michel Ferreira
DCC-FC & LIACC, University of Porto
`michel AT dcc.fc.up.pt`

New Publications

the title of the publication

(1)

the abstract/description goes here

the title of the next publication

(2)

the abstract/description goes here

Bibliography

[1] L.M. Herborg, M.G. Bentley, A.S. Clare, S.P. Rushton (2003) The spread of the Chinese mitten crab (Eriocheir sinensis) in Europe; the predictive value of an historical data set. *Hydrobiologia* 503: 21-28.

[2] P.W.W. Lurz, S.P. Rushton, L.A. Wauters, I. Currado, P. Mazzoglio, M.D.F. Shirley (2001) Predicting grey squirrel expansion in North Italy: a spatially explicit modelling approach. *Landscape Ecology* 16: 407-420.

Recent and Upcoming Events

the title of the event

a possible subtitle of the event

City, country, date

the text goes here

Your Name
Your organization
yourhomepage
mail AT mail net