Using DL for a Case-Based Explanation System

J.-L. Metzger, F. Le Ber, A. Napoli
LORIA - B.P. 239
54506 Vandœuvre-lès-Nancy
France
Email: {metzger, leber, napoli}@loria.fr

Abstract

This paper presents a knowledge-based system for land use interpretation and prediction. We describe our needs for representing knowledge and data, and for reasoning. We explain our choices : case-based reasoning within the framework of the description logic system RACER. Then, we present the knowledge base and the data we are working with. Data about spatial entities are represented as graphs and represented in the DL system accordingly. An example of graph manipulation is used to illustrate our purpose. Then, we propose a first synthesis of our research work and present an extension of the DL system necessary for going further.

1 Motivation

We are developing a knowledge-based system aimed at helping agronomists to analyze the management of farm territory, and to propose models about the evolution of farm territories and farmer practices. From the agronomist point of view, there are relationships between the farm management and the farm spatial organization. One of our goals is to build a system, named ROSA for Reasoning on Organization of Space in Agriculture, based on this hypothesis.

Section 2 describes the architecture we plan to use for the system ROSA and explains why we have chosen the description logic system RACER [7]. Section 3 presents the real-world data on which we are working, while section 4 describes how knowledge and data about farms are used for reasoning. In conclusion we point out the extensions on which we are working for fulfilling our needs.

2 A Case-Based Explanation System

Our research work mainly relies on case-based reasoning [11] and qualitative spatial reasoning [15] and especially reasoning with description logics (DL) [6, 14]. Case-based reasoning (CBR) is based on the use of past experiences called cases that are pairs *(problem, solution)* [12]. Our objective is to provide an explanation about the farm functioning while knowing its spatial organization. Thus, we consider that a problem is a farm management, a solution is an explanation on farm management. The explanation is based on the knowledge about the spatial organization of the farm.

A CBR system aims at solving a *target problem* denoted by `tgt` by means of a *case base* which is a finite set of cases. A case from the case base is called *source case* and denoted by `srce-case = (srce, Sol(srce))`. A *source problem* `srce` is a problem such that `(srce, Sol(srce))` belongs to the case base, i.e. it is a problem for which a solution `Sol(srce)` is known. CBR is based on three main operations: retrieval, adaptation and storage. The goal of retrieval is to find a case `srce-case` in the case base similar to the target problem `tgt`. Adaptation uses this retrieved case `srce-case` in order to build a solution `Sol(tgt)` to `tgt`. If the new case is of interest, it is stored in the case base.

After an early work on the use of object-based representation system and description logics system [13], we have decided to choose a DL system and especially the RACER system [7]. The expressivity of RACER, its well-defined semantics, the capability of dealing with concepts as well as with individuals and the efficiency of the reasoner are the main reasons of our choice. DLs have been used previously in CBR [5, 8, 9]. All these works argue on the capabilities of DLs with respect to the need of CBR systems. With their formal semantics and their ability to classify concepts and to recognize instances, DLs are well suited for managing knowledge and cases bases.

The knowledge base of our system is composed of domain knowledge and cases. Domain knowledge is used for enhancing the CBR operations. In particular, in our CBR system, we want to reason on domain knowledge and cases for classifying spatial structures and spatial relations [10]. Like in [5] or [8], domain knowledge is modeled in the TBox, while cases are described as individuals in the ABox. In RACER, we can take advantage of consistency checking (for the TBox) and classification (for querying the ABox).

The architecture of our system is shown in figure 1. It is based on an interface layer and a reasoning layer. The interface is designed to manage bases, for introducing new spatial description of farms and for displaying results. The reasoning layer is based on RACER to retrieve cases and adapt explanations.

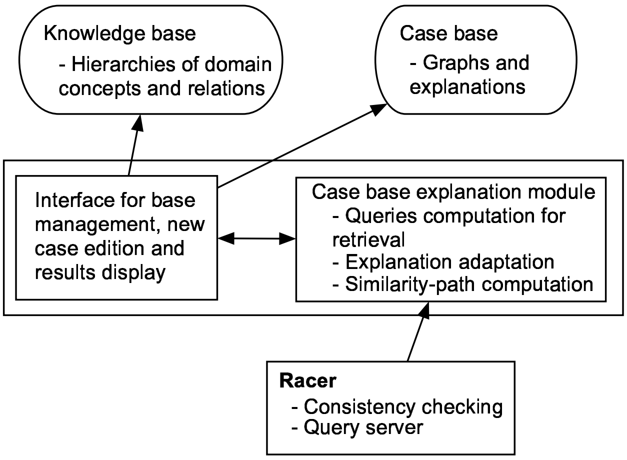

Figure 1: Architecture of our system ROSA.

3 Data and Domain Knowledge

In our system, the knowledge base contains the description of concepts about agronomy, land use and spatial relations. Concepts are organized in a hierarchy that is partially shown in figure 2.

Graphs are the base of our model for representing cases of farm spatial organization and farm functioning knowledge. Moreover, they are well-suited for representing complex real-world objects and for information sharing with domain experts and farmers. Our model is also inspired from the conceptual graph theory [3].

Graphs are composed of vertices and edges. Our graphs are built with two kinds of vertices : entities and links. An entity is an agronomic spatial object that relies on a concept. It can be qualified by different attributes. A link is a reification of a spatial relation. Like entities, links can be qualified by different attributes. Edges that connect entities to links are labeled by the qualification of entities in links. One graph represents the spatial organization about one farm. The farm functioning is described by explanations associated with parts of the graph. Hence, the explanation about a farm can be seen as a "sum" of all explanations associated to the graph.

The figure 3 shows two examples of parts of graphs from two instances of real-world graphs describing spatial organizations. These two graphs will be used in this article to explain our work.

For example, the explanation associated with the first graph `GrA` of the figure 3 is :

> *"the farmer has let a strip of grass (the meadow) to protect cereals from humidity, shadow and wild animals induced by the wood."*

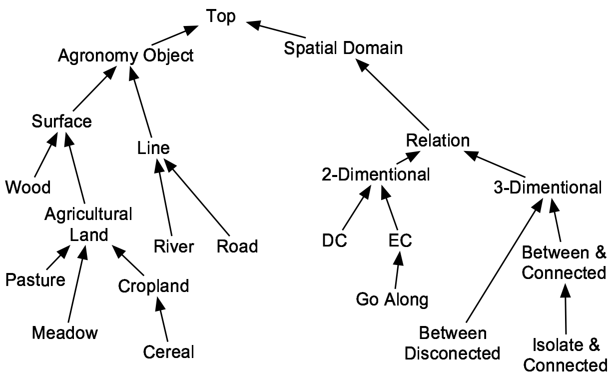

Figure 2: Hierarchy of concepts divided in two sub-hierarchies issued from the agronomy domain and the spatial knowledge.

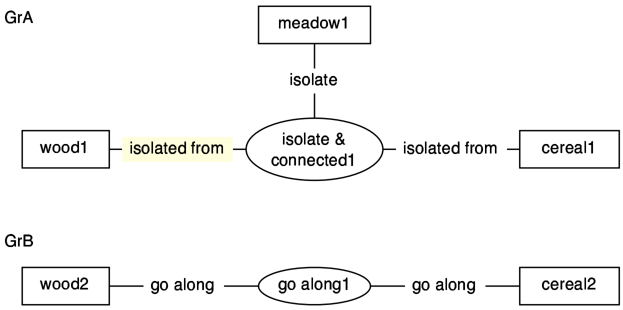

Figure 3: graphs representing some spatial organization of two farms A and B.

With the second graph `GrB` the explanation associated is:

> *"The wood is a constraint because there are humidity and shadow that are not good for cereals growth."*

Finally a case is a graph that describes part of the spatial organization of a farm associated with an explanation about its functioning.

4 Reasoning about spatial organization

Actually, explanations associated with a graph is given in natural language sentences by agronomists. Thus, when a new explanation has to be attached to a graph, it has to be coherent with explanations attached with graphs of the same kind. The validation of explanation is done by comparing different farms for extracting their differences and similarities concerning spatial organization.

To explain a new case, i.e. a given new farm described by its spatial organization in a graph, the CBR system tries to associate explanations to parts of

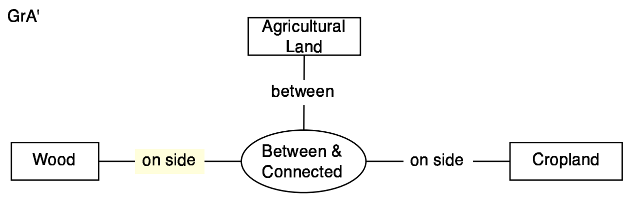

Figure 4: New graph `GrA'` obtained by modification of the first one in figure 3.

the graph. This is done with the help of a number of cases describing particular farms whose functioning is well understood.

The previous operations are based on the capabilities of comparing graphs. Two graphs, to be compared, have to be transformed in another one, i.e. a kind of "least common subsumer" of the two graphs [4, 1]. Three operations are allowed for transforming graphs : deletion, addition and substitution of vertices. Deletion or addition are based on specific rules of transformation based on domain knowledge. These rules are mostly given by the knowledge on relations according to the kind of vertices linked together. The substitution of a vertex is based on the classification mechanism : an instance can be substituted by the concept it belongs to; a concept can be substituted by its direct subsumer (in our case, at present, the concept of an instance and the subsumer of a concept are unique).

The "path" storing the transformation operations from one graph to another is a *similarity path*, used in case-base reasoning [12]. The similarity is computed with the help of an *edition distance* [2], giving the sum of costs of each transformations applied. Edition distance and similarity path are well adapted to our needs since they can be used in the adaptation process. [5] proposes this approach, but we need to extend it to deal with graphs.

For example, we describe the comparison of the two graphs of the figure 3. These graphs are described in the ABox. First, each vertex is classified according to the concepts hierarchy. Then, these graphs are transformed by generalization of their vertices. In the graph `GrA`, the vertex `meadow1` can be substituted by the concept `Meadow` and then by `Agricultural-Land`. In the same way, the concept of the vertex `cereal1` is `Cereal` and can be changed by `Cropland`. The relation `isolate&connected1` can be substituted by `Isolate&connected` and then by the `Between&Connected` relation. Finally, the instance `wood1` is subsituted by the concept `Wood` and the graph `GrA` is generalized into the graph `GrA'`, as shown figure 4.

In this particular case, we also know that `cereal1` and `meadow1` belong to the same farm, and that they are connected. Then, `cereal1` (generalized into `Cropland`) and `meadow1` (generalized into `Agricultural-Land`) are merge into a single vertex of type `Agricultural-Land`. Finally, beacause of the spatial

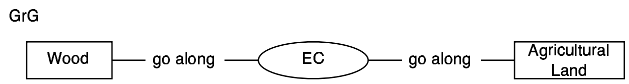

Figure 5: New graph GrG of the graph GrA' obtain by topological inference.

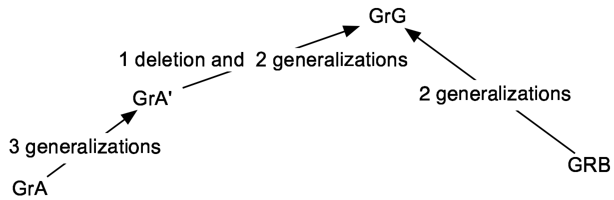

Figure 6: The similarity-path between graphs GrA and GrB.

position of meadow1 and wood1, it can be deduced that the EC relation hold between these two instances. Thus, the graph GrA' can be transformed again with three other modifications (two substitutions and a deletion) to obtain the more general graph GrG shown in figure 5.

The second graph GrB of the figure 3 can be transformed into a more general one by substituting cereal2 with the more general concept Agricultural-Land, by substituting wood2 with the concept Wood and by changing the relation go-Along1 to EC. The result is the same graph GrG described in figure 5.

In this example, the sequence of operations used to transform the graphs GrA and GrB into a more general graph GrG defines a similarity-path with seven transformations (figure 6). For each transformation, a cost based on the edition distance between the two graphs can be defined. The longer is the distance between the two graphs, more work has to be done for adapting explanations. Finally, the agronomists validate the fact that an adapted explanation makes sense or not.

5 Conclusion

We have given in this paper a first presentation of a system aimed at giving explanations on farm functioning for helping agronomists in the interpretation of land use. The system is based both on classification and case-based reasoning. There are a knowledge base and a case base. We have chosen the RACER DL for representing knowledge and reasoning. RACER DL is well suited for the construction of complex queries to retrieve instances. Reasoning mechanisms based on classification and consistency checking are powerful tools for the maintenance of the knowledge base. However more work remains to be done, especially with the graph transformations to ensure the existence to a least common subsumer of two graphs.

Our aim is to develop functionalities for the manipulation of graphs in RACER. For the moment we can deal with relations, i.e. a small graph with three or four vertices. But we need to compare larger graphs and define a classification mechanism for graphs. Such a functionality could be reused in many applications concerning spatial reasoning and other real-world tasks.

References

[1] F. Baader, R. Küsters, and R. Molitor. Computing least common subsumers in description logics with existential restrictions. In T. Dean, editor, *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI'99)*, pages 96–101. Morgan Kaufmann, 1999.

[2] H. Bunke and B.T. Messmer. Similarity measures for structured representations. In M.M. Richter, S. Wess, K.-D. Althoff, and F. Maurer, editors, *ewcbr93*, pages 26–31, 1993.

[3] M. Chein and M.-L. Mugnier. Conceptual graphs: Fundamental notions. *Revue d'intelligence artificielle*, 6(4):365–406, 1992.

[4] William W. Cohen, Alex Borgida, and Haym Hirsh. Computing least common subsumers in description logics. In Paul Rosenbloom and Peter Szolovits, editors, *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 754–761, Menlo Park, California, 1993. AAAI Press.

[5] P. A. González-Calero, B. Díaz-Agudo, and M. Gómez-Albarrán. Applying DLs for retrieval in case-based reasoning. In *Description Logics*, pages 51–55, july 1999.

[6] V. Haarslev, C. Lutz, and R. Möller. Foundations of spatioterminological reasoning with description logics. In A.G. Cohn, L. Schubert, and S.C. Shapiro, editors, *KR'98: Principles of Knowledge Representation and Reasoning*, pages 112–123. Morgan Kaufmann, San Francisco, California, 1998.

[7] V. Haarslev and R. Möller. Racer system description. In R. Goré, A. Leitsch, and T. Nipkow, editors, *IJCAR*, volume 2083 of *Lecture Notes in Computer Science*, pages 701–706, Siena, Italy, june 18-23 2001. Springer.

[8] G. Kamp. Using description logics for knowledge intensive case-based reasoning. In B. Faltings and I. Smith, editors, *Third European Workshop on Case-Based Reasoning (EWCBR'96), Lausanne, Switzerland*, pages 204–218. Springer-Verlag, Berlin, 1996.

[9] Jana Koehler. An application of terminological logics to case-based reasoning. In J. Doyle, E. Sandewall, and P. Torasso, editors, *KR'94: Principles of Knowledge Representation and Reasoning*, pages 351–362. Morgan Kaufmann, 1994.

[10] F. Le Ber, L. Mangelinck, and A. Napoli. Design and composition of lattices of topological relations for spatial representation and reasoning. Rapport de Recherche RR 4321, INRIA, 2001.

[11] D. B. Leake, editor. *Case-Based Reasoning. Experiences, Lessons and Future Direction*. MIT Press, 1996.

[12] J. Lieber and A. Napoli. Correct and complete retrieval for case-based problem-solving. In *ECAI'98*, pages 68–72, 1998.

[13] J.-L. Metzger. Acquisition, modélisation et représentation de connaissances pour le raisonnement à partir de cas : Application à l'étude d'organisations spatiales agricoles. Mémoire de dea informatique, Université Henri Poincaré - Nancy I, juillet 2000.

[14] R. Möller and M. Wessel. Terminological default reasoning about spatial information: A first step. In *Spatial Information Theory*, pages 189–204, 1999.

[15] L. Vieu. *In Spatial and Temporal Reasoning*, chapter Spatial representation and reasoning in artificial intelligence, pages 5–41. Kluwer Academic, 1997.