Journal of Universal Computer Science, vol. 9, no. 9 (2003), 1073-1095
submitted: 10/2/03, accepted: 9/6/03, appeared: 28/9/03 © J.UCS

Modeling and Comparing Farm Maps using Graphs and Case-based Reasoning

Florence Le Ber
(ENGEES, 1 quai Koch, BP 1039, 67070 Strasbourg CEDEX, France
& LORIA UMR 7503, BP 239, 54506 Vandœuvre-lès-Nancy, France
fleber@engees.u-strasbg.fr)

Amedeo Napoli
(LORIA UMR 7503, BP 239, 54506 Vandœuvre-lès-Nancy, France
napoli@loria.fr)

Jean-Luc Metzger
(LORIA UMR 7503, BP 239, 54506 Vandœuvre-lès-Nancy, France
metzger@loria.fr)

Sylvie Lardon
(ENGREF, 24 avenue des Landais, 63170 Aubière, France
lardon@engref.fr)

Abstract: In this paper, we present the knowledge-based system ROSA working on spatial and functional organizations in agriculture. The reasoning in ROSA combines hierarchical classification, case-based reasoning, and qualitative spatial reasoning. The goal of the system is twofold: formalizing and building a case base holding on farm spatial and functional organizations, and helping the analysis of new cases. Domain knowledge and cases are modeled with the help of the so-called *spatial organization graphs* (SOGs), and represented within a description logic system. Hierarchical case-based reasoning, involving classification and qualitative spatial reasoning, is used to compare and explain farm spatial structures modeled by SOGs. An example of case retrieval is proposed, followed by a global discussion on case-based reasoning in the ROSA system and related work.

Key Words: case-based reasoning, hierarchical classification, graphs, description logics, spatial structure analysis, agronomy.

Category: I.2.1 I.2.4

1 Introduction

In this paper, we present a knowledge-based system, called ROSA for *Reasoning about Organization of Space in Agriculture*, currently under development in collaboration with agronomists. The reasoning in the ROSA system follows the principles of case-based reasoning (CBR), where previously solved problems and their solutions, called cases, are used for solving new problems. The underlying assumption of CBR is that *similar* problems have similar solutions or similar problem-solving methods [19, 1]. In our research work, CBR relies on the agronomic assumption that there exists a strong relation between the spatial and the functional organizations of farms, and thus, that similar spatial organizations correspond to similar functional organizations. According to this assumption, and given a set of previously studied farm cases, the ROSA system has to help agronomists to analyze new problems holding on land use and land management in farms. Actually, the goal of our research work on this system is twofold: formalizing and building a case base on farm spatial and functional organizations with regard to environmental questions, and helping the modeling and analysis of new cases. Furthermore, original research results on spatial knowledge representation and reasoning arise from this study, and are presented hereafter.

In a first step of the present work, a model of the domain knowledge has been proposed, in accordance with agronomists. This model is based on *spatial organization graphs*, or SOGs, with labeled vertices and edges. Relying on these spatial organization graphs, farm *spatio-functional cases* have been designed: they mainly consist of a description of the land use, and an associated explanation linking spatial and functional organizations.

In a second step, the SOGs and the cases have been represented within a knowledge representation formalism, namely the description logic (DL) system RACER [15]. In this way, reasoning in the ROSA system relies on an original combination of hierarchical classification (in the description logic sense), case-based reasoning and qualitative spatial reasoning. In particular, spatial transformation rules are used for building *similarity paths* between SOGs. These paths are used in the case-based reasoning mechanism for comparing problems and adapting the solution from a source case to a new target problem [24].

The paper is organized as follows. The second and third parts present the context of this study and the modeling problem. The fourth part holds on spatial knowledge representation within description logics. We detail the case-based reasoning process in the fifth part and give an example of case retrieval in the sixth part. Finally, we discuss the present work, give a comparison with related works, and conclude.

2 The agronomic context

Agronomists of INRA-SAD[1] analyze and model land use and farm practices to answer environment and land management problems [10]. They perform inquiries

[1] INRA is the French research institute for agriculture and agronomy. SAD is a research department dealing with farm systems and rural development.

in various regions and study both farm spatial and functional organizations. They especially study two types of environment problems: in the Causses region (south of France), they focus on scrub invasion[2]; in Lorraine (north-east of France), the problem is water quality. These two types of problems are closely related to land use and more precisely to the spatial and functional organizations of farm territories. For example, in the Causses region, scrubs invade rough grazings if the grazing pressure of the ewe herd or the direct actions of the farmer are not sufficient. In Lorraine, water pollution is linked to corn, that is used for cattle feeding and that is cropped mostly on draining limestone plateaus.

For acquiring knowledge on the relationship between spatial and functional organizations of farm territories, agronomists have conducted farm inquiries. They have collected several information pieces like farm technical, economical, historical and geographical data. They have used the information given by the farmers and their own knowledge to produce synthetic maps of farm territories that express both the spatial and functional organizations of the considered farms. These maps are called *farm choreme*, because they rely on the *elementary choremes* formalized by geographers for modeling land organization and dynamics [5]. Elementary choremes are used as a guideline to recognize the principles of a farm spatial organization. Farm choremes are used to help comparisons, diagnosis, and management propositions on land use [20].

A farm choreme describes the spatial organization of fields, buildings, roads involved in land use and land management. It shows spatial structures that are interpreted by agronomists with respect to the farm functioning [29]. For example, Figure 1 shows a farm choreme modeling a farm in Causse Méjan, in the Causses region. The Causse Méjan is a high plateau (between 800 and 1200 m high) used by extensive sheep farms for milk or meat production. The territory of farm A is about 500 ha, and its ewe herd is about 400 animals with two lambings a year, in spring and in autumn.

The farm choreme of Figure 1 synthesizes the following inquiry informations.

- The farm territory is made of two connected allotments, denoted by **a1** and **a2**; the first one is centered on the farm house and the sheep pen, with an easy access to the main road; the second one is farther and linked to the sheep pen with a farm path.
- The areas near the sheep pen are used in spring and autumn by high need animals (i.e. ewes that have just lambed): a small crop field (**c1**), two blocks[3] of "nougats" (**b1**, **b2**), that are paddocks with small fields inside. The animal travel is minimized and the watch over is easy.
- The "nougat" (**n1**) is rather used in autumn when the farmer has time to

[2] The areas invaded by scrubs are progressively abandoned by the animals and lost for the farmer.

[3] A block is a set of crop fields, or paddocks, etc., that are close and managed in the same way.

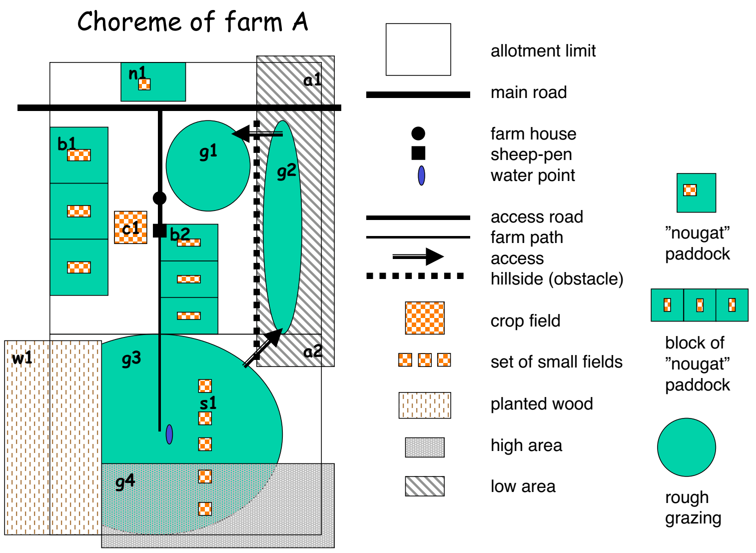

Figure 1: A choreme modeling a farm in Causse Méjan (south of France).

lead the animals across the road.

– The rough grazing denoted by **g1** is near the farm house, bordered by the road and the coast and thus easy to watch over. Since the sheep pen is near, animals do not go away.
– The rough grazing denoted by **g2** is used only when there is no more grass in the rough grazings denoted by **g3** and **g4**. Then the herd goes to rough grazing **g2** where grass remains green in summer (this area is shady and woody). The ewes naturally go back to the sheep pen through the rough grazing **g1** (the inverse way is not natural).
– Rough grazings **g3** and **g4** are farther and used in summer to maintain animals (lambs are weaned, ewes are still not mated). To reach the rough grazings, the farmer guides the herd to the farm path, and the ewes then go to the rough grazings **g3** and **g4**. Thanks to the water point, the herd can sometimes stay outside at night, otherwise it goes back to the sheep pen every day. The wood (**w1**) shelters the herd from summer heat. The small crop fields can be opened to pull the herd towards the rough grazing borders and thus spread the grazing pressure over the whole area.

To summarize, the farm territory is ex-centered, the paddocks near the sheep pen are used homogeneously during almost all the pasture season while the rough grazings are less used (*wrt* their surface) and only in summer. To better control scrub invasion on the farm territory, the paddocks near the sheep pen could be used in a more precise way, and the rough grazings should be cut into small paddocks encircling the water point. In this way, using the small fields and the wood, the grazing pressure should be easier to manage and to spread over the whole surface.

3 Modeling spatial and functional organizations

According to the objectives of the ROSA system, we have to define *spatio-functional cases*, that are spatial structures associated with functional explanations. That for, we rely on the synthetic information of the choremes, and computer scientists and agronomists have worked together to transform the choremes into spatial organization graphs. This work has lead to three main results. Firstly, we have specified the concepts used by the agronomists to describe farm spatial and functional organizations. Secondly, we have defined a set of farm spatio-functional cases. Finally, the transformation of choremes into graphs has, in turn, led the agronomists to improve the graphical representation of farm choremes [6, 21].

In this section, we detail the domain model of farm spatial and functional organizations, and we give examples of farm spatio-functional cases.

3.1 A model of domain knowledge

As mentioned in the previous section, the agronomists use several terms for describing spatial and functional aspects of farm organizations, such as:

- *land use*: crop fields, paddocks, rough grazings, temporary meadows, etc.
- *buildings and farm equipments*: farm house, sheep pen, water point, etc.
- *morpho-geological types*: plateau, coast, low and high areas, etc.
- *livestock*: lambs, ewes, dairy cows, etc.
- *farm functioning*: lambing, feeding, grazing management, etc.
- *spatial and spatio-functional relations*: border, near, far, separate, lead, etc.

The same term may cover various concepts, depending on the considered farm system or the region. We have thus defined two hierarchical domain models, one for Lorraine and one for the Causses region, that include the description of the elements listed above, and that can be considered as domain ontologies. For example, Figure 2 shows a part of the Causses domain model, denoted by $\mathcal{H}_{\mathcal{CC}}$. Spatial regions are clustered into three categories, i.e. surfaces, lines and points, involving specific spatial properties, e.g. a point may be inside a surface but not the converse. Points correspond to buildings and equipments. Lines correspond to roads, paths and rivers. Surfaces are categorized with respect to the land use: for example rough grazings, “nougats” and paddocks are grasslands, while crop fields, “almonds” (i.e. crops surrounded with a grass band) and temporary meadows are arable lands.

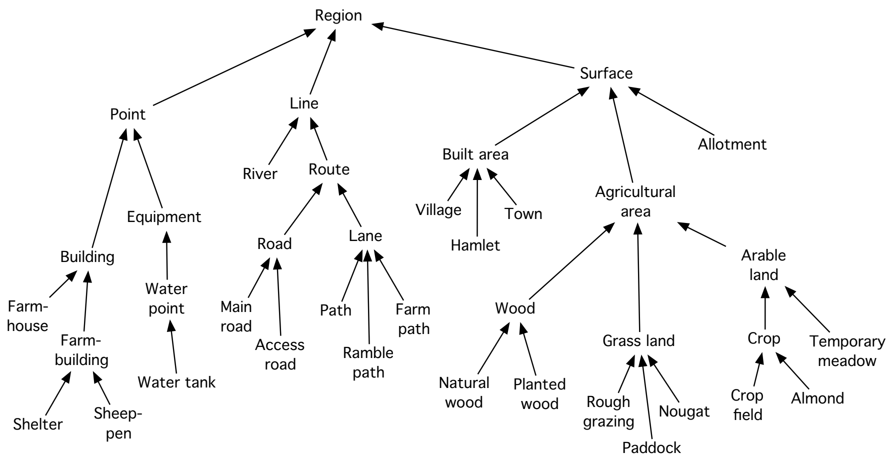

Figure 2: The $\mathcal{H}_{\mathcal{CC}}$ domain model for the Causses region (partial view: land use, buildings and equipements).

A third specific hierarchical model, denoted by $\mathcal{H}_{\mathcal{CR}}$, has been designed for spatial relations. This hierarchical model relies on qualitative models of spatial relations proposed in [7, 16, 33]. The relations describing farm spatial organizations are organized according to the usual three categories, topology, distance, and orientation, within the $\mathcal{H}_{\mathcal{CR}}$ hierarchy. For topological relations, we rely on the axiomatisation of mereotopology based on the connection and convex-hull primitives, that include the following relations: *proper part* (PP), *contains as a proper part* (PP-1), *identical* (EQ), *partially overlaps* (PO), *disconnected* (DC), *externally connected* (EC), *inside*, *p-inside*, *outside*, etc. [33, 9]. Regarding qualitative distance, we define two granularity levels as proposed in [7]: the first level includes three categories, namely *near*, *medium*, and *far*; the second level includes four categories, namely *very-near*, *medium-near*, *medium-far* and *very-far*. Regarding orientation only the *between* relation and some specializations (between with or without connection) are considered. Orientation and distance relations can be axiomatised in conjunction with mereotopology by adding a *sphere* primitive, as proposed in [30, 3]. The hierarchy of relations $\mathcal{H}_{\mathcal{CR}}$ is partially described in Figure 3.

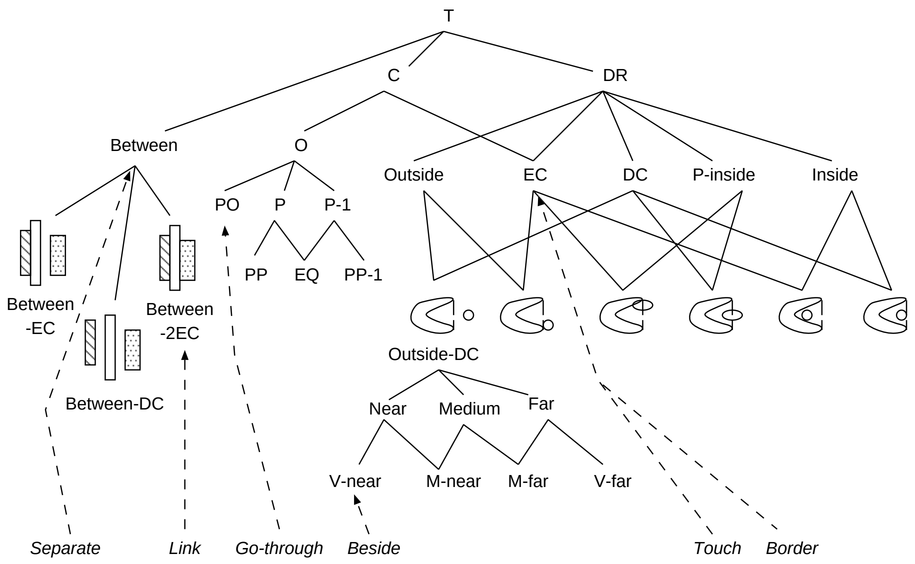

Figure 3: The spatial relation hierarchy $\mathcal{H}_{\mathcal{CR}}$, inspired by [31]. Spatio-functional relations are related to spatial relations (dashed lines).

Moreover, the spatio-functional relations used by the agronomists to describe farm organizations are clustered with respect to the $\mathcal{H}_{\mathcal{CR}}$ hierarchy. For example, *touch* and *border* are EC relations, *go-through* is a PO relation, while *beside* is a *very-near* relation, and *separate* is a *between* relation. It must be noticed that spatio-functional relations are described in the domain model of the ROSA system, but spatial reasoning is performed on the spatial relations only, as shown below.

Furthermore, since our objective is to manipulate and to compare structures, we have defined a set of pragmatic spatial transformation rules. Transformations rules are of different types. Some of them are based on the neighborhood of spatial relations in $\mathcal{H}_{\mathcal{CR}}$, as it is done for **RCC-8** base relations in [33, 8]. For instance the two relations EC and DC are neighbors, as well as EC and PO, or *very-near* and *medium-near*. Transformation rules also include composition of distance and topological relations as discussed in [7, 12, 32]. Finally, we have designed specific transformation rules that rely both on spatial relations and on characteristics of spatial regions.

For example, we have introduced the following rules, which apply to specific spatial relations (see Figure 4):

– Tr-1 (Inside,Near): *if a region* `A` *is inside a region* `B` *that is near a region* `C`*,* *then* `A` *is near* `C`*.*

– Tr-2 (Between-2EC): *if a surface* `B` *is between a surface* `A` *and a surface* `C`*, where* `A` *is connected to* `B` *and* `C` *is connected to* `B`*, and if* `A` *and* `B` *can be unified into a unique surface* `A+B`*, then* `A+B` *is externally connected with* `C`*.*
– Tr-3 (Between-2EC): *if a line* `L` *is between and connects a surface* `A` *and a surface* `C`*, and if* `L` *can be removed, then* `A` *and* `C` *are disconnected.*

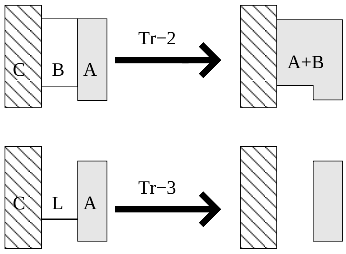

Figure 4: Spatial transformation rules Tr-2 and Tr-3.

It must be noticed that our objective is not to define a complete table and an axiomatic model of composition on $\mathcal{H}_{\mathcal{CR}}$ relations – this work remains to be done – but instead, to specify a number of pragmatic and working rules for comparing farm spatial organizations, within the cbr process (see section 5).

3.2 Farm spatio-functional cases

A spatio-functional case corresponds to the description of a spatial structure and the associated functional interpretation. It is used as a reference for interpreting farm spatial organizations. Actually, a case is restricted to a part of a particular farm: in this way, a farm is described by several cases, at various scales. A case is modeled by an *explained graph*, denoted by e-sog, associating a sog with an *explanation*. An explanation is a textual statement (plain text), that can be rather complex, holding on the functional interpretation of the current spatial structure. A sog is a bipartite graph composed of labeled vertices and edges. The label of a vertex denotes a spatial entity (in $\mathcal{H}_{\mathcal{CC}}$) or a spatial relation (in $\mathcal{H}_{\mathcal{CR}}$). A spatial entity is always related to a spatial relation and reciprocally. The edges are labeled with terms referring to the role played by the spatial entities within the relations, mainly *subject* or *object*.

Figure 5 shows three cases associated with the farm depicted in Figure 1. The e-sog denoted by **A** is a model of a particular spatial structure with an associated explanation meaning that this particular crop field, which is near the farm house and has a high production potential, is used by animals (mainly lambing ewes in spring) needing close watch and energy nutriments. The E-SOG **A** is composed of five vertices: `sheep-pen`, `farm-house` and `crop-field` refer to spatial entities while `near` and `beside` refer to spatial relations. The roles of the entities are given by the labels associated with the edges between the vertices, namely `subject` (the relations are symmetrical).

The E-SOG denoted by **B** describes another spatial structure with another explanation which means that the access to the farm is easy since it is linked to the main road by a smaller road. The explanation of the E-SOG denoted by **C** means that since the block of nougat paddocks is near the sheep pen, it can be used in sequence (starting from the sheep pen and going farther as the season passes) by high need animals in spring. The small fields inside the nougat paddocks are used to attract the animals farther.

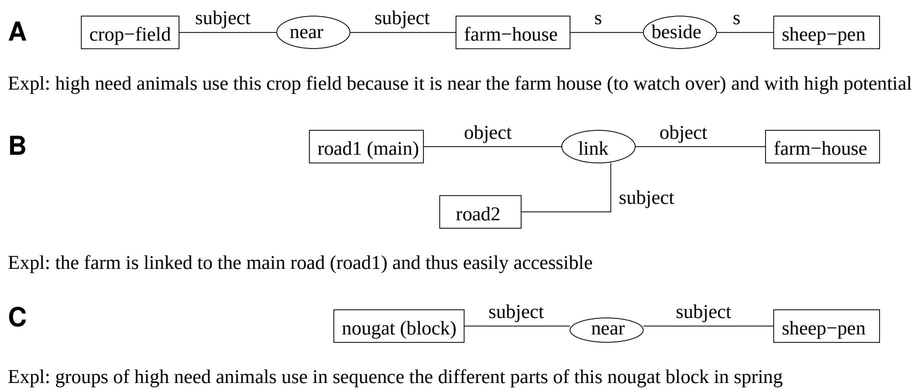

Figure 5: Three cases representing a part of the farm of Figure 1.

These examples show that a farm organization is modeled with several E-SOGs. The different E-SOGs can be combined into a general SOG describing the whole structure of a farm, and the explanations can be combined accordingly. The set of E-SOGs considered as reference cases for the analysis of new farms defines the *farm case base* of the ROSA system.

4 Knowledge Representation

The objective of the ROSA system is to help the analysis of farm spatial and functional organizations, *wrt* a set of previously analyzed farms. The reasoning in the ROSA system relies mainly on classification and case-based reasoning (CBR).

Accordingly, the architecture of the ROSA system is composed of the following elements (Figure 6):

– a domain knowledge base, including the three hierarchies of agricultural and spatial concepts,
– a case base, including a set of E-SOGs,
– a CBR module combining a number of reasoning mechanisms, for spatial reasoning, graph matching, similarity computing and explanation adaptation,
– an interface for introducing the SOGs and E-SOGs, and displaying the results.

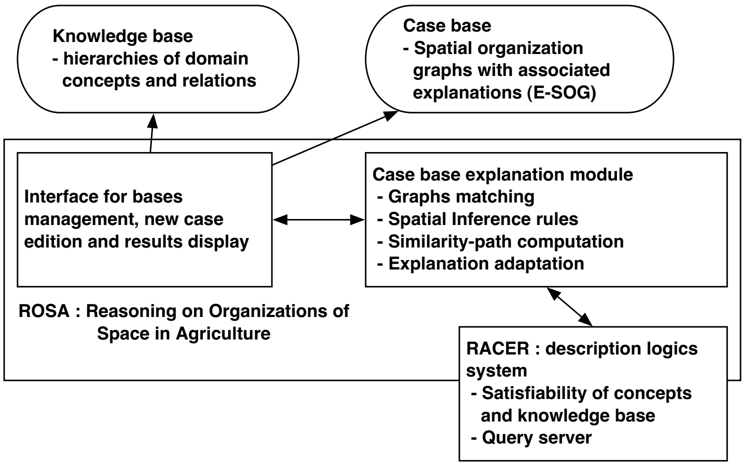

Figure 6: The architecture of the ROSA system.

The representation and the reasoning mechanisms in the CBR module rely on the RACER description logic system [15], that provides an efficient subsomption test and an associated classification procedure. The domain knowledge and the cases are represented within the RACER system. Below, we first briefly introduce the framework of description logics, and then we detail the representation of domain knowledge and cases.

4.1 Description logics

The present research work is a continuation of a previous research work on the representation of spatial structures and topological relations [26, 22, 23]. The knowledge representation formalism formerly used was relying on an object-based representation system, with representation primitives similar to those of description logics. In the present work, the knowledge representation formalism has to enable the representation and the manipulation of complex spatial structures modeled by SOGs and E-SOGs. Thus, the use of a description logic system has been a natural choice for extending and improving the previous research results. Furthermore, description logics have proven to be valuable and efficient for spatial reasoning [13, 14].

Briefly (see for example [11, 2] for a complete survey), a DL system allows the representation of knowledge using descriptions that can be concepts, roles or individuals. A concept represents a set of individuals and is composed of roles representing properties of the concept and relations with other concepts. Descriptions of concepts and roles are built according to a set of constructors (e.g. `and`, `or`, `not`,...). Concepts can be either primitive or defined, and are organized within a hierarchy using the subsumption relation. The concept hierarchy (also called `TBox`) corresponds to the ontological level of knowledge representation. Beside the concept hierarchy, a set of assertions in which individuals are involved forms the so-called `ABox` of the DL system. An assertion can be a concept or a role instantiation. Reasoning is based on concept satisfiability, knowledge base (`TBox` + `ABox`) satisfiability, classification and instantiation. In the framework of the ROSA system, the two main procedures are classification and instantiation. The former is used to classify a concept in the concept hierarchy according to the search of its most specific subsumers and most general subsumees. The latter is used for finding the concepts of which a given individual may be an instance.

We have chosen to use the RACER DL system [15] because it provides a very rich set of constructors for descriptions, and the means for exploiting a concept hierarchy and a set of assertions. Actually, RACER implements the $\mathcal{SHIQ}$ description logic [14], one of the most expressive DL at the moment. The subsumption test (on which is based classification) and the instantiation test are efficient operations in RACER. Furthermore, concrete domain (number, string) are also available, as well as an interface with XML.

4.2 The domain knowledge and case bases in the ROSA system

The `TBox` of the ROSA system is composed of a hierarchy of concepts denoted by $\mathcal{H}_\mathcal{C}$, a hierarchy of indexes denoted by $\mathcal{H}_{Idx}$ (see below), and a hierarchy of roles denoted by $\mathcal{H}_\mathcal{R}$. At present, the hierarchy $\mathcal{H}_\mathcal{R}$ only includes the roles `object` and `subject`. The hierarchy $\mathcal{H}_\mathcal{C}$ includes the domain hierarchies $\mathcal{H}_{\mathcal{CC}}$ and $\mathcal{H}_{\mathcal{CR}}$ and introduces a concept `Sog`[4] representing the SOGs in a general way, as a list of vertices whose type[5] is the concept `Vertex`. A vertex may be a spatial entity whose type is the concept `Region`, or a spatial relation between entities

[4] In the following, the name of concepts is capitalized while the name of individuals is in normal size.

[5] The *type* of a concept is itself; the type of an individual `i` is the most specific concept of which `i` is an instance (this concept is unique in the ROSA system).

whose type is the concept `Relation`. It can be noticed here that only a list of vertices is recorded for a `Sog`; edges are retrieved when necessary through the vertices of type `Relation` to which they are attached. The concept `E-Sog`, is a specialization of the concept `Sog` and encloses an explanation that (at present) is a string of type `Explanation`.

A particular SOG is represented as an instance of the concepts `Sog` or `E-Sog`. For example, the E-SOG denoted by **A** on Figure 5 is represented as an instance of `E-Sog` with:

- a list of vertices constituted by instances of `Crop-field`, `Farm-house`, and `Sheep-pen` linked to each other by an instance of the relation `Near` and an instance of `V-near` (`Beside` is a specialization of `V-near`, see Figure 3);
- an instance of the concept `Explanation` representing the text of the explanation given in Figure 5.

At present, the case base of the ROSA system is made of the complete descriptions of nine farms (four from Causses and five from Lorraine). The description of a farm relies on an average of ten cases represented by instances of `E-Sog`. The cases are clustered and organized within the hierarchy $\mathcal{H}_{Idx}$, with respect to generic SOGs, called *indexes* (Figure 7). Finally, the elements in the domain knowledge base and in the case base (`TBox` + `ABox`) provide the knowledge on farm spatial and functional organizations used for solving new analysis problems, as this is explained in the next section.

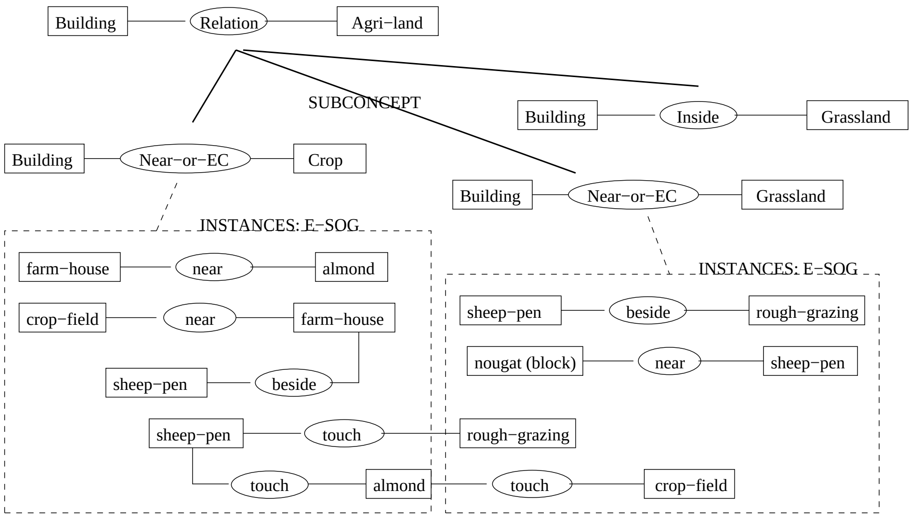

Figure 7: The hierarchical indexed case base $\mathcal{H}_{Idx}$ (explanations are not represented). The instance at the bottom of the figure depends on two concepts.

5 Reasoning in the ROSA system

In the ROSA system, the analysis of a new farm is based on a hierarchical CBR process. Recall that CBR relies on three main operations : retrieval, adaptation and memorization. Given a new *target* problem to be solved, denoted by `tgt`, the system searches in the case base for a *source* case `(srce,Sol(srce))` where the `srce` problem is similar to the `tgt` problem. The solution of the `tgt` problem is built by adapting the solution `Sol(srce)` of the source problem. Finally, the new pair `(tgt,Sol(tgt))` may be memorized in the case base according to its interest.

In our framework, a problem statement corresponds to a query for the analysis of the spatial structure of a new farm, and the solution is a set of explanations on the farm functional organization. Actually, a target problem is represented as a SOG, say `tgt`, for which an explanation `Expl(tgt)` must be found. A source case is an E-SOG `(srce,Expl(srce))`, where `srce` is a SOG and `Expl(srce)` the corresponding explanation. The retrieval operation consists in searching for one or more source cases `(srce,Expl(srce))` providing an explanation for the `tgt` problem. In practice, the retrieval operation consists in an exploration of the ROSA case base for finding a source case matching (a part of) the `tgt` problem. The case retrieval is based on the *strong* and *smooth classification* principles introduced in [24, 25]. The objective is to define a so-called *similarity path* between the problems `tgt` and `srce`, that can be used for adapting the `Expl(srce)` explanation into a new explanation, `Expl(tgt)`. Below, the retrieval process is described, following an introduction of strong and smooth classifications in the ROSA system.

5.1 CBR principles in ROSA

The retrieval operation in the ROSA system is based on a classification of the current problem statement, denoted by `tgt`, in the index hierarchy $\mathcal{H}_{Idx}$ of the case base. It determines one or more source cases, that can be reused for solving the target problem after adaptation. When one or more source cases are available, then a source case, denoted by `srce`, is selected according to a given preference criterion, and the adaptation process is activated: this corresponds to the strong classification operation. When no source case is available, then the smooth classification process has to be activated.

The adaptation operation is linked to the retrieval operation by the notion of *similarity path*. Such a path can be seen as a sequence of operations, i.e. generalization and specialization, linking the `srce` problem statement of the source case to the `tgt` problem statement. In this way, the strong classification process is used to design the following similarity path:

$$\texttt{srce} \sqsubseteq \texttt{idx(srce)} \sqsupseteq \texttt{tgt}$$

The relation $\texttt{srce} \sqsubseteq \texttt{idx(srce)}$ means that the index associated with the source case, i.e. `idx(srce)`, is more general than the statement `srce` of the source case[6]. The relation $\texttt{idx(srce)} \sqsupseteq \texttt{tgt}$ means that the index of the source case is more general than the target problem. The similarity path is valid only if the index `idx(srce)` is different from the top of the index hierarchy $\mathcal{H}_{Idx}$.

As soon as a similarity path has been found, the adaptation operation may be performed, based on two main operations, generalization and specialization. The first step within adaptation is to generalize the explanation `Expl(srce)` to produce an explanation `Expl(idx(srce))` that in turn is specialized into an explanation `Expl(tgt)`. The generalization and specialization operations used for building the explanation `Expl(tgt)` are parallel to the corresponding operations in the similarity path.

$$\begin{array}{ccccc} \texttt{srce} & \sqsubseteq & \texttt{idx(srce)} & \sqsupseteq & \texttt{tgt} \\ \downarrow & & & & \uparrow \\ \texttt{Expl(srce)} & \sqsubseteq & \texttt{Expl(idx(srce))} & \sqsupseteq & \texttt{Expl(tgt)} \end{array}$$

The strong classification process relies on an *exact matching* between the source and the target problem statements. This is not always true: it is then necessary to *transform* the target problem in order that the strong classification operation may be applied[7]. In other words: *smooth classification = strong classification + transformation*. In this way, the similarity path being built has the following format:

$$\texttt{srce} \sqsubseteq \texttt{idx(srce)} \sqsupseteq \textsc{Tr}\texttt{(tgt)} \leftarrow \texttt{tgt}$$

The transformations TR being applied to `tgt` in the ROSA system are spatial transformation rules, as described in Section 3.1. In a more general way, these transformations depends on the application domain. According to this smooth similarity path, the adaptation is based on three main operations, *generalization*, *specialization*, and *transformation*. The generalization and specialization operations are performed as in strong classification, and a transformation operation is applied to the `Expl(TR(tgt))` explanation, in parallel with the transformation operation in the similarity path, to produce an explanation for the `tgt` problem.

$$\begin{array}{ccccccc} \texttt{srce} & \sqsubseteq & \texttt{idx(srce)} & \sqsupseteq & \textsc{Tr}\texttt{(tgt)} & \leftarrow & \texttt{tgt} \\ \downarrow & & & & & & \uparrow \\ \texttt{Expl(srce)} & \sqsubseteq & \texttt{Expl(idx(srce))} & \sqsupseteq & \texttt{Expl(}\textsc{Tr}\texttt{(tgt))} & \rightarrow & \texttt{Expl(tgt)} \end{array}$$

A cost is associated to a similarity path, depending on the operations that are used to build the different steps of the path: generalization, specialization

[6] As an index must be.

[7] In [24, 25], both the source and the target problem statements are transformed within the smooth classification operation.

and transformation. One important thing, that will not be discussed here, is to find a similarity path of minimal cost, with the goal of minimizing the number of adaptation operations.

5.2 Case retrieval strategy

We now detail the retrieval process in the context of the ROSA system. Let `tgt` be an instance of `Sog`, representing the spatial organization of a new farm to be analyzed. The search for a source case that can explain the `tgt` problem is based on strong classification, possibly followed by smooth classification.

5.2.1 Strong classification

Given `tgt`, an instance of `Sog`, we call `Tgt` the type of `tgt`, i.e. the concept of which `tgt` is an instance. The concept `Tgt` is classified within the index hierarchy $\mathcal{H}_{Idx}$ of the case base. This classification process determines the set of the most specific subsumers of `Tgt`, denoted by $S_{Idx}(\texttt{Tgt})$[8]. An index `G` in $S_{Idx}(\texttt{Tgt})$ corresponds to a part of the target problem: actually the graph `G` matches a subgraph of `Tgt`.

According to the hirarchical organization of the case base $\mathcal{H}_{Idx}$, `G` is an index of a set of cases S_G. The following similarity path is built between each `srce` case of S_G and the `tgt` problem.

$$\texttt{srce} \sqsubseteq \texttt{idx(srce) = G} \sqsupseteq \texttt{Tgt} \sqsupseteq \texttt{tgt}$$

In this way, each element of $S_{Idx}(\texttt{Tgt})$ leads to a set of cases that are associated to different parts of `tgt`. A corresponding number of similarity paths are built, and a set of explanations can be obtained accordingly.

5.2.2 Smooth classification

Some parts of `Tgt` may remain unanalyzed after the strong classification process, and are undertaken within the smooth classification process. The parts that have been analyzed within the strong classification process are marked and are no longer taken into account unless it is explicitly stated by the system user. However, any part of `Tgt`, even already analyzed, can be selected for an alternative study based on the smooth classification process.

The smooth classification process relies on two main operations. Firstly, a transformation rule TR is applied to the relation vertices of the unanalyzed parts of `Tgt`, provided that the triggering conditions of TR are fulfilled. For example, the rule TR-2 can be applied to a relation vertex if its type is `Between-2EC` and

[8] If $\texttt{G} \in S_{Idx}(\texttt{Tgt})$, there is no element $\texttt{G'} \neq \texttt{G}$ in $\mathcal{H}_{Idx}$ such that $\texttt{Tgt} \sqsubseteq \texttt{G'} \sqsubseteq \texttt{G}$.

if the neighboring region vertices are of `Surface` type. Only one transformation rule is allowed to be applied to a relation vertex (composition of rules is not allowed). The choice of the rule to be applied is guided by the system user.

Then, the transformed SOG, say `Tr(Tgt)`, is classified into the index hierarchy $\mathcal{H}_{Idx}$ of the case base. The process is carried on in the same way as in the strong classification process, producing similarity paths between the source cases and the `tgt` problem, as follows :

$$\texttt{srce} \sqsubseteq \texttt{idx(srce)} \sqsupseteq \texttt{Tr(Tgt)} \leftarrow \texttt{Tgt} \sqsupseteq \texttt{tgt}$$

Finally, the result of the smooth classification process is a set of source cases and similarity paths, associated to particular parts of `tgt`.

6 Example

In this section we describe the analysis of a `tgt` SOG representing a part of the spatial organization of a farm in Causse Méjan. The `tgt` SOG is described in Figure 8, with the two source cases and the index retrieved during the strong classification process. According to the classification of `Tgt`, i.e. the concept instantiating `tgt`, within the index hierarchy $\mathcal{H}_{Idx}$, the most specific subsumer of `Tgt` is the index `Index1` that matches the subgraph (`Paddock`,`V-near`,`Sheep-pen`) of `Tgt` (bold, see Figure 8). Then `Tgt` is compared to the instances of `Index1`, namely `srce11` and `srce12`. The comparison is based on vertex matching and the similarity paths between the target and source SOGs are designed accordingly:

- from `Tgt` to `srce11`, the following operations have to be done:
  - `Paddock - rough-grazing` : generalization of `Paddock` into `Grassland` and then specialization into `Rough-grazing` (`Paddock` $\sqsubseteq$ `Grassland` $\sqsupseteq$ `Rough-grazing`, see Figure 2).
  - `V-near - beside`: `Beside` is specialization of `V-near`.
  - `Sheep-pen - sheep-pen`: `sheep-pen` is an instance of `Sheep-pen`.
- from `Tgt` to `srce12`, the following operations have to be done:
  - generalization of `V-near` into `Near` (`V-near` $\sqsubseteq$ `Near`, see Figure 3),
  - generalization of `Paddock` into `Grassland` and then specialization into `Nougat` (`Paddock` $\sqsubseteq$ `Grassland` $\sqsupseteq$ `Nougat`).

Then, as some parts of `Tgt` have not been analyzed, the smooth classification process is activated. Transformation rules are applied to the relation vertices that have not been classified, i.e. `EC`, `Inside`, `Far`. For example, the rule Tr-1 (see section 3.1) may be applied, according to a user request. Actually, in the `Tgt` SOG, the `Crop-field` vertex is linked to the `Paddock` vertex through an `Inside` vertex, and the `Paddock` vertex is linked to the `Sheep-pen` vertex through a `V-near` vertex. When applying the rule Tr-1 to the vertices `Inside`

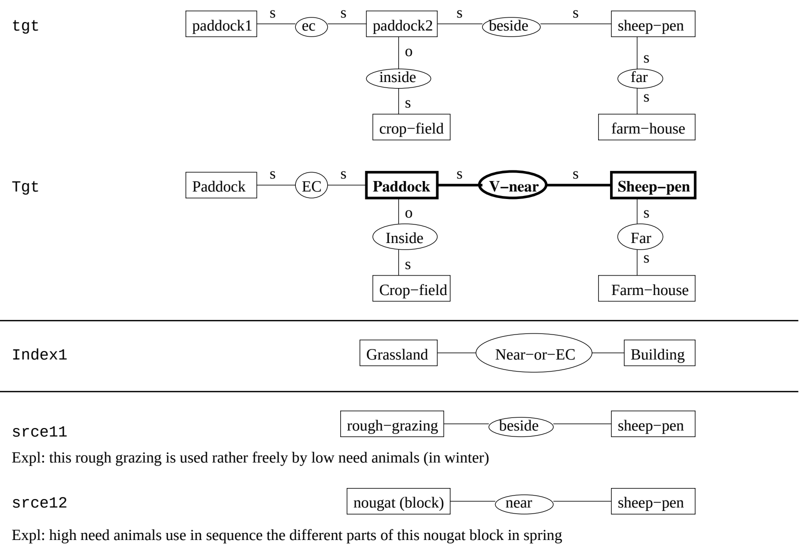

Figure 8: A simplified example of case retrieval within the strong classification process. Recall that `Beside` is a specialization of `V-near`.

and `V-near`, a new graph is obtained, where the `Crop-field` and `Sheep-pen` vertices are linked through a new `Near` vertex (bold, see Figure 9).

The `Tr-1(Tgt)` SOG resulting from the application of Tr-1 to `Tgt` is classified within the index hierarchy $\mathcal{H}_{Idx}$. The index `Index2` described in Figure 9 is a most specific subsumer of `Tr-1(Tgt)`, as it matches the subgraph (`Crop-field`,`Near`, `Sheep-pen`) of `Tr-1(Tgt)`. Similarity paths between the transformed SOG and the source cases that are instances of `Index2`, namely `srce21` and `srce22`, are computed as follows:

– from `Tr-1(Tgt)` to `srce21`, the following operations have to be performed:
  - generalization of `Sheep-pen` into `Building` and then specialization into `Farm-house` (`Sheep-pen` $\sqsubseteq$ `Farm-building` $\sqsubseteq$ `Building` $\sqsupseteq$ `Farm-house`, see Figure 2),
  - generalization of `Crop-field` into `Crop` and then specialization into `Almond` (`Crop-field` $\sqsubseteq$ `Crop` $\sqsupseteq$ `Almond`).

– from `Tr-1(Tgt)` to `srce22`, the following operations have to be done:
  - generalization of `Sheep-pen` into `Building` and then specialization into `Farm-house` (just as before).

Finally the results of the retrieval process are presented to the agronomist for rejection or validation. In the present example, the agronomist has rejected the `srce11` source case, and validated the three other cases. The rejection of

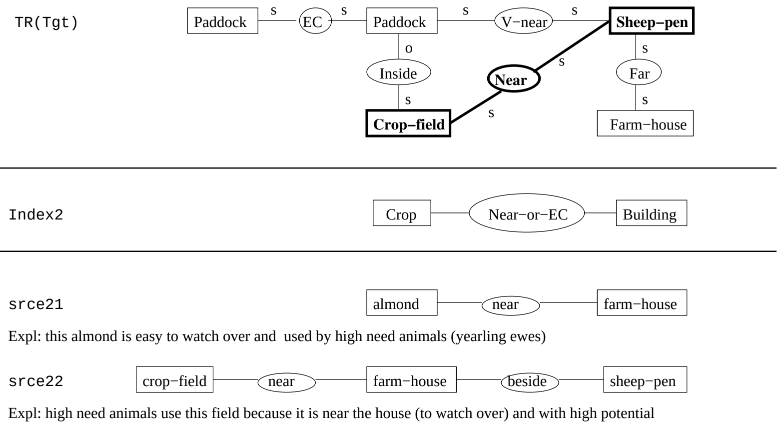

Figure 9: A simplified example of case retrieval within the smooth classification process.

`srce11` is motivated by the fact that rough grazing is used by low need animals, while a paddock beside a sheep pen is rather used by high need animals (as the other retrieved source cases can be interpreted).

Actually, experiments of this kind are meaningful and give directions for completing and revising the knowledge bases, and for improving reasoning and problem solving.

7 Discussion and related work

7.1 Spatial representation and description logics

The work presented here follows a previous work on the representation and the classification of spatial structures in an object-based knowledge representation system [23]. In this previous work, the problem of recognizing spatial structures on images is considered as a classification problem, where patterns of structures are represented by classes and the image regions are represented by individuals to be classified with respect to classes. Topological relations are reified, i.e. represented as classes with attributes and facets, and organized within a lattice-based hierarchy. In the present work, we are mainly interested in the functional interpretation of spatial structures. Thus, instead of relying on predefined structure patterns, the reasoning in the ROSA system takes advantage of a collection of individuals, namely E-SOGs, that are used as references for analyzing new spatial structures. Classification-based reasoning is completed by case-based reasoning, and both are combined within the reasoning process of the ROSA system.

In addition, in the present work, no procedural module is needed for recognition purposes, and thus the choice of a description logic system is well founded. Indeed, DL systems have been used in an efficient way in a number of research works on qualitative spatial reasoning. In [13, 14], the authors propose to represent spatial objects and relations within the RACER system, and to perform spatial reasoning based on consistency checking and classification. From the application point of view, regions are represented by polygons (elements of the concrete domain), and reasoning is carried on relations between polygons, for recognizing regions with specific characteristics and answering queries to a map database. In [27], spatio-temporal default reasoning is introduced: default knowledge is represented within rules and used for completing and making more precise queries to a map database. In our case, transformation rules are used on spatial structures for a better matching with the individuals of a spatial case base.

7.2 The combination of classification and case-based reasoning

Two fundamental modes of reasoning are used in the ROSA system: classification of concepts and relations, and hierarchical case-based reasoning for the functional analysis of spatial structures. Regarding CBR, our work is mainly inspired by similarity paths introduced in [24, 25], and can also be related to a number of other works [18, 17, 34]. In [18], the cases are indexed by a hierarchy of concepts; the retrieval and adaptation operations are based on the classification of concepts. The decomposition of adaptation is based on two kinds of classification, and can be likened to strong and smooth classification. Moreover, some of the ideas on memorization in [18] could be reused in our own context. In [34], a case is described by an individual and the similarity between a source case and a target case is based on the search of the least common subsumer (LCS) of two concepts. The source cases retrieved are then classified according to a dissimilarity measure between the source and the target cases. In our approach, the cases are also represented by individuals, and the matching between a source case and a target case is based on the search of a graph playing a role similar to the LCS concept. In addition, transformation operations such as insertion, deletion and substitution, are used to build a similarity path.

Regarding graph classification, the main inspiration comes from the works described in [28, 35]. Indeed, the matching of spatial structures presents similar characteristics to the matching of molecular structures. In the present work, the matching process relies on the comparison of the *composition* of structures, and is considered from two viewpoints.

– When the considered SOG is reduced to a triple `(region,relation,region)` – or a quadruple in case of ternary relations –, then the classification mechanism in RACER is sufficient on its own. Actually SOGs containing only one

relation vertex are represented as defined concepts in RACER, and can be directly classified within the index hierarchy $\mathcal{H}_{Idx}$.
- When the considered SOG is composed of at least two relations, then an external module for graph matching has to be invoked. This external module has been especially designed for testing SOG matching, and is associated to the RACER system.

7.3 The adaptation problem

A number of problems have to be solved for providing an efficient and generic adaptation process in ROSA. As explained before, the adaptation process in CBR is mainly dependent on the matching of source and target problems. In the ROSA system, the matching of SOGs relies on concept classification and transformation rules. These rules play an important role in the design of similarity paths, and thus in the adaptation process. Furthermore other rules have to be designed to adapt the explanations from a source case to a target problem.

At present, the system returns one or more similarity paths leading from the retrieved source cases to the target problem. The agronomist is in charge of rejecting or validating the corresponding explanations. This decision process can also be undertaken by the system, at least in part, using adaptation knowledge (as discussed in [4, 19]), but this aspect is out of the scope of the present paper.

8 Conclusion

The present paper describes the knowledge-based system ROSA, that works on spatial and functional organizations of farms. The objective of this system is to help agronomists to analyze the spatial organization of farms with respect to their functioning. Our approach is based on classification and hierarchical case-based reasoning in a description logics framework, namely the RACER system. A hierarchical domain model and a hierarchical case base have been designed and implemented. Spatial structures are modeled within spatial organization graphs, and manipulated on the basis of classification procedures and transformation rules. The retrieval process in ROSA is based on graph matching and the definition of similarity paths between graphs. Moreover, the whole approach presented in this paper can be considered as being of general interest for the representation and manipulation of spatial structures as graphs.

The ROSA system is still under development, and there are a number of points that must be made more precise and worked further. One important point holds on the SOG matching process based on transformation rules. The choice of the rules is fundamental *wrt* the building of useful similarity paths for the adaptation step. Regarding this point, the following questions have still to be investigated: which graphs are comparable, and how adaptation can be performed accordingly, i.e. how explanations can be represented and adapted in the general case? To answer these questions, experiments must be carried on for designing and comparing farm choremes and SOGs.

References

1. A. Aamodt and E. Plaza. Case-Based reasoning: Foundational Issues, Methodological Variations, and System Approachs. *AICOM*, 7(1):39–59, 1994.
2. F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, editors. *The Description Logic Handbook Theory, Implementation and Applications*. Cambridge University Press, 2003.
3. B. Bennett, Anthony G. Cohn, Paolo Torrini, and Shyamanta M. Hazarika. A foundation for region-based qualitative geometry. In W. Horn, editor, *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI 2000), Berlin, Germany*, pages 204–208. John Wiley & Sons Ltd., 2000.
4. R. Bergmann and W. Wilke. Towards a New Formal Model of Transformational Adaptation in Case-Based Reasoning. In H. Prade, editor, *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI 98), Brighton, UK*, pages 53–57. John Wiley & Sons Ltd., 1998.
5. R. Brunet. La carte-modèle et les chorèmes. *Mappemonde*, 86(4):3–6, 1986. In French.
6. M. Capitaine, S. Lardon, F. Le Ber, and J.-L. Metzger. Chorèmes et graphes pour modéliser les interactions entre organisation spatiale et fonctionnement des exploitations agricoles. In T. Libourel, editor, *Géomatique et espace rural. Journées CASSINI 2001, Montpellier, France*, pages 145–163. SIGMA, 2001. In French.
7. E. Clementini, P. D. Felice, and D. Hernàndez. Qualitative Representation of positional information. *Artificial Intelligence*, 95:317–356, 1997.
8. A. G. Cohn, B. Bennett, J. Gooday, and N. M. Gotts. Representing and reasoning with qualitative spatial relations about regions. In O. Stock, editor, *Spatial and Temporal Reasoning*, pages 97–134. Kluwer Academic Publishers, 1997.
9. Z. Cui, A.G. Cohn, and D.A. Randell. Qualitative and Topological Relationships in Spatial Databases. In D. J. Abel and B. C. Ooi, editors, *Advances in Spatial Databases, Third International Symposium SSD'93, Singapore*, LNCS 692, pages 296–315. Springer-Verlag, 1993.
10. J.-P. Deffontaines, J.-P. Cheylan, S. Lardon, and H. Théry. Managing rural areas. From pratices to model. In J. Brossier, L. de Bonneval, and E. Landais, editors, *Systems studies in agriculture and rural development*, Science Update, pages 383–392. INRA Éditions, 1993.
11. F.-M. Donini, M. Lenzerini, D. Nardi, and A. Schaerf. Reasoning in description logics. In G. Brewka, editor, *Principles of Knowledge Representation*, pages 191–236. CSLI Publications, 1996.
12. M. J. Egehofer. Reasoning about binary topological relations. In O. Gunther and H.-J. Schek, editors, *Advances in Spatial Databases, Second International Symposium, SSD'91*, LNCS 525, pages 143–160. Springer, 1991.
13. V. Haarslev, C. Lutz, and R. Möller. Foundations of Spatioterminological Reasoning with Description Logics. In *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR 98), Trento, Italy*, pages 112–123. Morgan Kaufmann Publishers, 1998.
14. V. Haarslev, C. Lutz, and R. Möller. A description logic with concrete domains and a role-forming predicate operator. *Journal of Logic and Computation*, 9(3):351–384, 1999.
15. V. Haarslev, R. Möller, and A.-Y. Turhan. Racer user's guide and reference manual. Hamburg University, 2001.

16. D. Hernández. *Qualitative Representation of Spatial Knowledge.* LNAI 804. Springer-Verlag, 1994.
17. 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, 1996.
18. J. Koehler. Planning from second principles. *Artificial Intelligence*, 87:145–186, 1996.
19. J. Kolodner. *Case Based Reasoning.* Morgan Kaufman Publishers, 1993.
20. S. Lardon, M. Capitaine, and M. Benoît. Les modèles graphiques pour représenter l'organisation spatiale des activités agricoles. In *Représentations graphiques dans les systèmes complexes naturels et artificiels. Journées de Rochebrune, Megève, France*, pages 127–150. ENST, 2000. In French.
21. F. Le Ber, C. Brassac, and J.-L. Metzger. Analyse de l'interaction experts – informaticiens pour la modélisation de connaissances spatiales. In *IC'2002, Journées Francophones d'Ingénierie des Connaissances, Rouen, France*, pages 29–38. INSA Rouen, 2002. In French.
22. F. Le Ber and A. Napoli. Design and comparison of lattices of topological relations based on Galois lattice theory. In *Proceedings of the Eighth International Conference on Principles of Knowledge Representation and Reasoning (KR 2002), Toulouse, France*, pages 37–46. Morgan Kaufmann Publishers, 2002.
23. F. Le Ber and A. Napoli. The design of an object-based system for representing and classifying spatial structures and relations. *Journal of Universal Computer Science*, 8(8):751–773, 2002. Special Issue on Spatial and Temporal Reasoning.
24. J. Lieber and A. Napoli. Using Classification in Case-Based Planning. In W. Wahlster, editor, *Proceedings of the 12th European Conference on Artificial Intelligence (ECAI 96), Budapest, Hungary*, pages 132–136. John Wiley & Sons Ltd., 1996.
25. J. Lieber and A. Napoli. Correct and Complete Retrieval for Case-Based Problem-Solving. In H. Prade, editor, *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI 98), Brighton, UK*, pages 68–72. John Wiley & Sons Ltd., 1998.
26. L. Mangelinck. Représentation et classification de structures spatiales. Application à la reconnaissance de paysages agricoles. PhD Thesis, Université Henri Poincaré, Nancy, 1998. In French.
27. R. Möller and M. Wessel. Terminological Default Reasoning about spatial information: A First Step. In C. Freksa and D. M. Mark, editors, *Spatial Information Theory, Cognitive and Computational Foundations of Geographic Information Science, International Conference COSIT'99, Stade, Germany*, LNCS 1661, pages 189–204. Springer, 1999.
28. A. Napoli, C. Laurenço, and R. Ducournau. An object-based representation system for organic synthesis planning. *International Journal of Human-Computer Studies*, 41(1/2):5–32, 1994.
29. M. Naïtlho and S. Lardon. Representing spatial organisation in extensive livestock farming. In *Integrating Animal Science Advances into the Search of Sustainability, 5th Int. Livestock Farming Systems Symposium, Fribourg, Switzerland, August 1999*, pages 187–190, 2000.
30. D. A. Randell and A. G. Cohn. Modelling Topological and Metrical Properties in Physical Processes. In R. Brachman, H. Levesque, and R. Reiter, editors, *Conference on Principles of Knowledge Representation and Reasoning (KR 89), Toronto, Canada*, pages 55–66. Morgan Kaufmann Publishers, 1989.
31. D. A. Randell and A. G. Cohn. Exploiting Lattices in a Theory of Space and Time. *Computers & Mathematics With Applications*, 23(6-9):459–476, 1992.
32. D. A. Randell, A. G. Cohn, and Z. Cui. Computing Transitivity Tables: A Challenge For Automated Theorem Provers. In *Proceedings of the 11th Interna-*

tional Conference on Automated Deduction, CADE'92, LNCS 607, pages 786–790. Springer-Verlag, 1992.
33. D. A. Randell, Z. Cui, and A. G. Cohn. A Spatial Logic based on Regions and Connection. In *3rd International Conference on Principles of Knowledge Representation and Reasoning (KR 92)*, pages 165–176. Morgan Kaufmann Publishers, 1992.
34. S. Salotti and V. Ventos. Study and formalization of a case-based reasoning system using a description logic. In B. Smyth and P. Cunningham, editors, *Advances in Case-Based Reasoning, Proceedings of the 4th European Workshop (EWCBR-98), Dublin, Ireland*, LNCS 1488, pages 286–297. Springer, 1998.
35. P. Vismara. Reconnaissance et représentation d'éléments structuraux pour la description d'objets complexes. Application à l'élaboration de stratégies de synthèse en chimie organique. PhD Thesis, Université des Sciences et Techniques du Languedoc, Montpellier, 1995. In French.