

Journal of Visual Language and Computing

journal homepage: www.ksiresearch.org/jvlc/

A Mathematical Language for the Modeling of Geospatial Static Rules

Robert Laurini^a

^aKnowledge Systems Institute, USA, and University of Lyon, France

ARTICLE INFO

Article History:

Submitted 4.8.2019

Revised 8.8.2019

Second Revision 5.20.2019

Accepted 8.15.2019

Keywords:

Geographic knowledge

Geographic rules

Formal grammar

Smart cities

ABSTRACT

More and more in information technologies, rules are considered as first-class citizens; and many applications in business intelligence are built on rules. But in territorial intelligence and in smart city planning, few works have been done in this direction. The role of this paper is to show the importance of geographic rules and to propose, beyond the modeling in natural language, a mathematical language to model them. This language is primary based on logic and set theory in which some relations and operators coming from topology, computational geometry and operation research are integrated, and will cover only static rules. By mathematical language, one means that it is independent from any software product or application, and its formal grammar is presented by means of syntactic diagrams. Several examples are provided especially in urban planning.

© 2019 KSI Research

1. Introduction

According to Graham [6] and Morgan [14], rules must be considered as first-class citizens in information technologies, meaning first that several computer-based activities must be revisited. By definition, a rule is a sequence antecedents-implication-consequents which can be noted by $(A) \Rightarrow (B)$, in which A is a conjunction of conditions. Instead of antecedents, sometimes the expression ‘premise’ is used. In logic, B can be either a disjunction of conditions or a set of assertions. They come from the so-called Horn clauses which are also the basis of logic programming, where it is common to write definite clauses in the form of an implication: $(p \wedge q \wedge \dots \wedge t) \Rightarrow u$.

However, in urban planning, rules are essentially coming from laws and by-laws. Moreover, some experts can use other rules in their daily practice, sometimes

called best practices. In addition to that, more and more specialists in spatial data mining analytics can discover what they call associative rules.

In the knowledge society, in smart city management, it is important not only to identify rules, but also to combine them to automate reasoning. Facing this objective, the scope of this paper is to propose a language in order to model rules. Indeed, rules are often made explicit in natural language; two main sources will be used, namely rules written in natural language and associative rules as extracted from data mining. This language will be based on mathematics without taking into account practical implementation. In other words, this is not a computer language (code). Voluntarily, this paper will not deal with 3D issues, neither with temporal issues: only geographic static rules will be considered.

This paper is organized as follows. First, some elements will be given to explain the role of rules in computing, and especially in geoprocessing. Then the formal grammar of this mathematical language will be detailed. Finally, examples especially in urban planning, will clarify the expressive power of this language.

^aCorresponding author

Email address: Roberto.Laurini@gmail.com

Website: <http://www.laurini.net/robert/>

ORCID: 0000-0003-0426-4030

2. About rules in IT

In this section, the importance of rules will be rapidly emphasized in Business Intelligence. In Business Intelligence, generally, their implementation is based on two grammatical structures IF-THEN-Fact and IF-THEN-Action (Ross [16]). The first serves above all to involve new facts, that is new objects, attribute values, new relationships between objects. As to the second, it is to involve new actions. But who will be in charge of such new actions? In some cases, the computer itself may run procedures or send a message to other devices; in others, particularly in regulatory contexts, a decision maker (for example, the CEO of a company) must himself initiate the action. Another interpretation could be that the choice of alternatives of an action, for example when a law, in some well-defined contexts, opens many perspectives. Thus, a rule is a basic element of a strategy to build reasoning. In contrast to algorithms, they are expressed declaratively. Among business rules, Dietz [3] distinguishes between three categories:

- rejectors typically those related to quality control, that allow a rejection (rejection rules),
- producers such as those determining new values (ex VAT calculation); they can be considered as rules of production of information,
- and projectors such as those related to the replenishment of stocks.

To conclude this section, let us mention that lots of business applications are based on rules, and several computer languages for encoding business rules have been proposed, as XML extensions, such as SWRL (Espinasse, [5]) or RuleML (Boley [1], Boley et al. [2]). The simplest of those extensions is as follows:

```
<Implies>
  <if>
<..>
  </if>
  <then>
  <..>
  </then>
</Implies>
```

Also, very recently, a visual language has been proposed such as by Pittl et al. [15], based on SWRL.

3. Rules in geoprocessing

In contrast, concerning geographic sciences and their applications, few works have been done in the spirit of knowledge engineering. However, let us mention Malerba et al. [13], Salleb-Aouissi et al. [18] and Varadharajulu et al. [22].

From spatial data mining, for instance, Malerba et al. (2003) have discovered the following rule:

$$is_a(X, large_town) \wedge intersects(X, Y) \wedge is_a(Y, road) \rightarrow Intersects(X, Z) \wedge is_a(Z, road) \wedge Z \neq Y (91\%, 85\%).$$

Which states that 'If a large town X intersects a road Y, then X intersects a road Z distinct from Y with 91% support and 85% confidence'. In other words, one can say that *is_a* allows a kind of definition, *intersects* a relation and \neq an additional condition.

Salleb-Aouissi et al. [18] have studied geology and mineral deposits in South America. They extended the existential and universal quantifiers (\forall , \exists) by incorporating buffer zones (here 5 km), and have proposed rules such as:

$$Mines - \exists_{5k}^3 Faults \rightarrow True$$

$$Mines - \exists_{1k}^1 Volcano \rightarrow (active=yes)$$

The first rule means that there exist at least 3 faults within 5 km of a target object (mineral deposits), whereas the second states that there exists at least one active volcano within 1 km of a target object. They also propose to write $\forall_{5km}^{80\%}$ for considering 80% of items with 5 km. Remark that the distance is taken as a modification of the quantifiers. As it could be nice for colocation rules, other extensions must be defined for other relations coming from topology or computational geometry.

However, Varadharajulu et al. have proposed the following rule for checking road length against road type by using SWRL (remark that the length is given as a data, not computed from computational geometry); more exactly, a road less than 200 m long must be called a "close":

```
NEWROAD(?R1), ROADSUFFIX(?R1, ?T1),
hasLength(?R1, ?200), SameAs(?T1, ?Close)
-> isAllowed(?R1, true)
```

All those three examples illustrate the difficulties to incorporate issues coming from topology and computational geometry. Indeed, the governance of smart cities must be based on both artificial intelligence and collaborative human intelligence. New domains such as sensors networks, big data, deep learning, geovisualization, etc. must be mobilized together with knowledge engineering to provide more efficient systems for citizens. Do not forget that one of the scopes of urban big data is to discover novel patterns and rules.

3.1 About geographic rules

In Laurini et al. [10], several examples of geographic rules were given, and some important semantical aspects were extracted. Anyhow, geographic rules are a way to model what Shoorcheh [21] has called spatial causality.

3.2 Generalities about the model

Based on the previous descriptions, a general diagram for rules can be designed regrouping all aspects, their origin, components, temporal dimension, mathematic tools, their management, their usage and the various modes of implication (Figure 1).

- For physical rules, the implication is mandatory;
- For legal rules, the implication is also mandatory, but the sanctions may or may not exist; in this case, a second rule with the negative conditions will lead to the sanctions;
- For best practices, the implication is more or less a kind of recommendation; in other words, nobody is obliged to follow this kind of rules; perhaps some additional conditions could be considered; at a first approximation, random variates can help to select best practices if any.
- For rules coming from data mining, it is necessary to provide support and confidence. See Shekar et al. [20] for details.

Figure 1 explains the main characteristics of geospatial rules, namely, origin, mathematical tools, temporal dimension, semantics, modality, usage and management. Remark that an existing rule can be superseded by another novel rule, sometimes for a period or a narrower place.

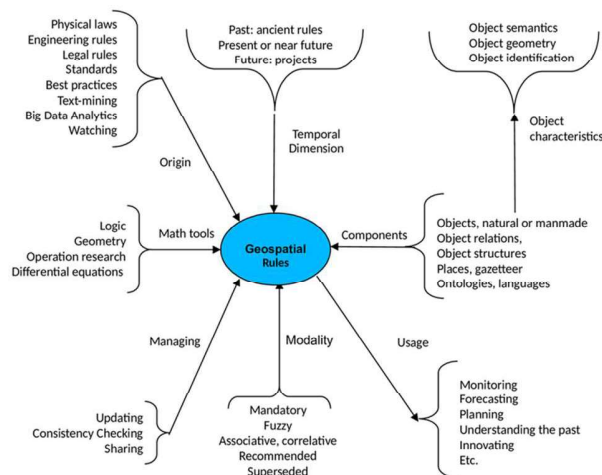


Fig. 1. Main characteristics of geospatial rules.

Now that basic components of geospatial rules are given, let us try to design a mathematical model.

4. A Mathematical Model

This model will be based not only on logics and set theory, but also on computational geometry, topology and fuzzy set theory. Before giving the grammar, some precisions must be done to explain the main components. Anyway, in this model, we assume that:

- a) a global geographic object called *Earth* includes all existing geographic objects (*GO*) and territories (*Terr*),
- b) an ontology will describe their types/classes (whatever is the concept) together with some specific attributes and generic relations between them,
- c) all objects will have valued attributes; in the case of new objects, the attributes are set to null,
- d) there exists a set named *Projects* which comprises all possible environmental and urban projects,
- e) there exists a gazetteer integrating all place names possibly with different variants and in different languages,
- f) it is assumed that all information is correct and consistent,
- g) it is assumed that there are no problems neither regarding geometry accuracy, nor multi-representation,
- h) there are no considerations for storing, implementation, optimization, etc.,
- i) this language is not a rigid language as in computer sciences, but a sketch in which everybody can add symbols, functions, etc.

4.1 Geographic sets and collections

The first elements are the *Earth* and the sets of Geographic Objects (*GO*). As some of them are well characterized (roads, buildings, islands, etc.), others need some clarifications about their definitions. The scope of this paper is not to contribute to those characterizations as works regarding geographic ontologies have revealed several difficulties of categorization. Smith and Varzi [19] have tried to clarify by distinguishing “Fiat” and “Bona Fide” boundaries. Many objects are known by their name (such Sahara), but the boundaries are not well defined, similarly for the Rocky Mountains.

4.2 About places

All places belong to *Earth* and can be described according to various solutions. The simplest ones are by a name, for instance Argentina, the second by a set of coordinates (polygons). Do not forget that a place can form a non-connected polygon (for instance, a country with its islands). More complex solutions can be defined, e.g. the group of countries in which people are driving on

the left, a city district enclosed in a set of streets.

4.3 About geographic objects

All geographic objects (*GO*) have three types of attributes, for identifying them, for describing them from a geometric point of view and the semantic attributes. The dot notation can be used, for instance "*A.population*". Concerning identification, sometimes ID's can be used, but it is the more common to use a place name or a toponym. As several places can have the same toponym (e.g. Mississippi), the location will resolve ambiguities. Concerning geometry, remember that sometimes an object can have several geometries (multi-representation) sometimes taken at different scales.

Semantic attributes can have different formats such as alphanumeric, Boolean, multimedia, etc. In some cases, fuzzy values can be used such as "near", "far", "low", "high", etc. See Kainz [7] for details and examples.

However, in urban and environmental planning, one needs to consider projects. Indeed, projects have different phases. First, they are designed (design phase) maybe with several sub-steps (preliminary, front-end, etc.). Then there is a legal phase (building permit) in which the project can be approved or rejected. The construction phase itself will followed, sometimes delayed for archaeological or financial reasons.

Finally, the projected object becomes really a geographic feature, finished or not. In other words, it is necessary to consider a *Project* superclass including any geographic object-to-be, for instance *Project.Road*. To simplify the problem, the ontology of those objects-to-be will similar to the ontology of existing geographic objects.

4.4 Relations, functions and procedures

Geographic relations are the basis of geographic reasoning. Those relations include Egenhofer et al. [4] 2D topological relations (Figure 2) and other relations.

Some of them are defined for types, and some other for specific geographic objects. Remember that a relation is Boolean (true or false) and can be checked in conditions.

Functions are also useful, some of them coming from computational geometry. To name a few, *Union*, *Intersection*, *Minus*, *Distance*, *Centroid*, *Buffer*, etc., and they will be invoked when necessary. Anyone can add additional functions.

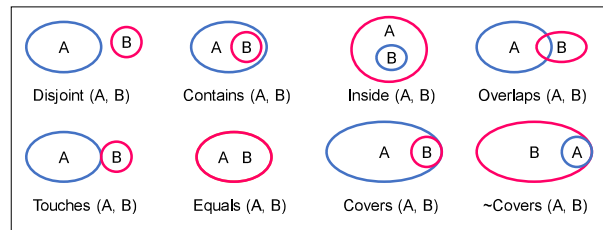


Fig. 2. 2D Topological relations. Egenhofer et al. [4].

Two functions will be extensively used, *Geom* for the geometry and *Topo* for names. By definition *Geom* (*A*) is equivalent to *A.Geom*, and *Topo* (*A*) to *A.Toponym*. Although they are not indispensable, they will be useful in some cases.

Moreover, regarding existing models, the best solution is to encapsulate those mathematical models into procedures for using them when necessary. In addition, a special procedure can be used to call another rule.

4.5 Semantics of Symbols

In this model, the mainstream meaning of mathematical symbol will be accepted. However, due to this special context, some semantics must be fixed for some symbols. Let us explain some problems.

4.5.1 Implications

Three types of implications must be considered. The symbol \Rightarrow will be used for logical assertions coming from physics or legal regulations. For best practices, the expression " \Rightarrow [BP]" will be used, and for associative rules coming from data mining, this symbol will be accompanied with the support and the confidence levels.

4.5.2 About equalities

Remember that the very common sign = has three meanings, (i) for defining something, (ii) for an assignment of value, and (iii) for comparisons. We will use the following symbols:

- a) = as a comparator in Boolean conditions; so that the answer will be either true or false,
- b) \equiv for definitions, especially of new variables,
- c) := for assigning a new value to a known variable.

4.5.3 The More, the Merrier

Often rules are written according to the style "the more of this, the more of that", or "the less of this, the more of that"; for instance: "the more of traffic, the more of pollution". To solve this problem, three solutions are possible.

- To compare two geographic points – A_1 and A_2 – and to show the evolution in space of some attributes such as if $A_1 > A_2$ then $F(A_1) > F(A_2)$, or $G(A_1) < G(A_2)$,
- A solution can be to write $+x \Rightarrow -y$, the semantics of which are when x increases, then y diminishes; for instance, when saying, the higher, the colder, this sentence can be roughly encoded by $+elevation \Rightarrow -temperature$,
- If a function between both variables is known, it can be used directly.

4.5.4 About homologies

When one needs to compare two geometric shapes, it is commonly accepted that there are some point coordinates which are slightly different. Moreover, for the same object, one person has measured 100 points, whereas another 500. Sometimes, the symbol \approx is used for comparing numerical values. But when we need to compare types and toponyms of geographic objects, other semantics must be ascertained. Indeed, for strings of characters, the Levenshtein distance [12] is used. For instance, between Iraq and Iran the Levenshtein distance is 1, whereas they are very different geographic objects. Regarding types, one can consider that the concepts *street* and *road* are similar. To solve this problem, I have proposed to use the symbol \bowtie (Laurini, [9, 11]) for homologies. So that, considering two geographic objects, we can write:

- $Geom(A) \bowtie Geom(B)$ for comparing these geometric shapes (hence this is equivalent to $Geom(A) \approx Geom(B)$).
- "London" \bowtie "Londres" (London, UK, is called Londres, both in French and Spanish languages; another example is "Washington D.C." \bowtie "District of Columbia").
- $Type(A) \bowtie Type(B)$ when the concept names are different but are corresponding to the same concept, maybe coming from different ontologies.

4.5.5 Other symbols

In logic, the entailment symbol \vDash is used to denote an affirmation taking the context into account. By writing $\vDash Contains(A, B)$, one claims that this relationship must be considered as True. Whereas by writing only $Contains(A, B)$, one declares that he is considering this relationship, but he has no hints whether it is True or False. Regarding the other symbols coming from the set theory ($\forall, \exists, \cap, \cup, \supset, \subset, \in, \vee, \wedge, \oplus$, etc.), they have their common meaning.

4.5.6 Consequents

In logic, remember that a conjunctive list of antecedents implies a disjunctive list of consequents so that $A_1 \wedge A_2 \wedge A_3 \wedge \dots A_i \Rightarrow C_1 \vee C_2 \vee C_3 \vee \dots C_j$ in which both A

and C are Boolean expressions. In the disjunctive list, the semantics are not clear: are all C 's true or only a subset? Moreover, in our case, the consequents are not always Boolean (remember IF-THEN-Fact and IF-THEN-Action). To solve this problem, several decisions must be made: (i) discarding the symbol \vee , and separating consequents by the symbol “;”, meaning that all of them are independent and must be enabled; (ii) to use set bracket parentheses $\{\}$ to delimit several consequents if any.

4.5.7 About complex places

Two points of view can now be considered, according to the set theory and according to topology: we can write either $Ghana \in Earth$ or $Contains(Earth, Ghana)$. By extension, the set of left-driving countries can be defined by either $(UK, Ireland, Japan, \text{etc.}) \in LD_Countries$, or $Geom(LD_Countries) \equiv Union(Geom(UK), Geom(Ireland), Geom(Japan), \text{etc.})$. More generally, to define a place, all computational geometry functions can be used. Concerning streets, it could be important to delimit only a sector, for instance by giving civic numbers; for instance, $Geom(S) \equiv StreetSector(Street_name, Civic_number1, Civic_number2)$.

To delimit a city district by a set of surrounding streets, a special function can be invoked $SurroundedByStreet(Street1, Street2, Street3, Street4, \text{etc.})$.

4.5.8 Remarks

Here are some additional remarks. First, we need to split antecedents in context and conditions. Indeed, in the previous examples, several things were not explicit, i.e. the considered objects and sets were hinted or hidden: those aspects will be described in the context, maybe including definitions and constraints. Moreover, if one needs to consider a negative rule, only the conditions are denied, but not the constraints.

Secondly, let us enlighten the differences between $a \in b$ and $contains(c, d)$. In those expressions, b is a set, and a belongs to this set in the sense of the set theory, whereas c and d are geometric objects and d is inside c from a topological point of view.

5. Formalism

Any geographic rule will be designed as pictured in Figure 3, containing namely Context, Conditions, Implication and Consequents with the following separators : and \blacksquare . Context will describe the main variables; Conditions, one or more Boolean criteria to

follow; Implication, the modality; and Consequents, all possible consequences that are to be run.

The grammar is presented by means of the Railroad Diagram Generator (Rademacher, [17]) which is an interesting way to present formally the syntactic structure of a language.

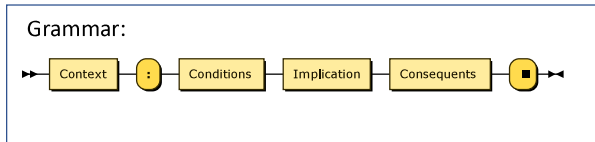


Fig. 3. Formal definition of the grammar.

The Context will be defined as usual from the set theory (Figure 4) by defining variables and their sets, possibly followed by Definitions.

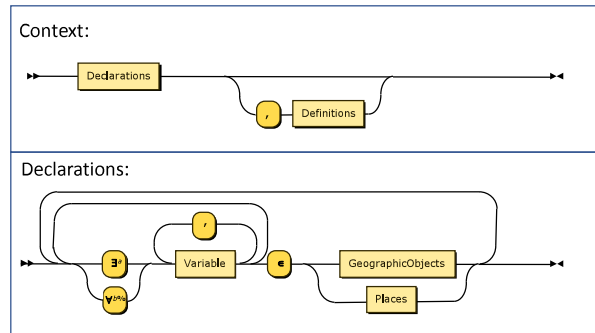


Fig. 4. Context and Declarations.

In Figure 4, as previously explained the extended quantifiers (Salleb-Aoussi et al. [18]) will be used in which a stands either for void (i.e. 1) or for a number greater than 1, whereas b is a percentage. By definition, $\exists^1 = \exists$ and $\forall^{100\%} = \forall$.

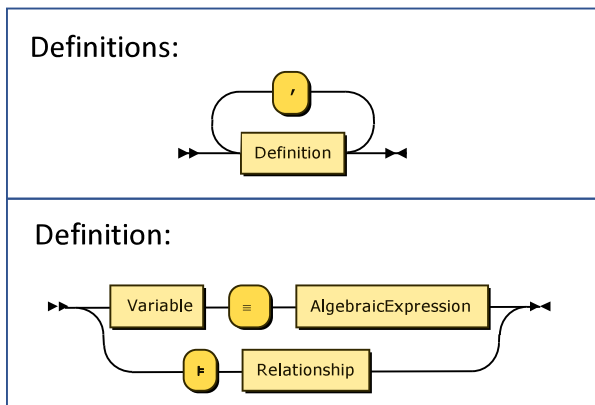


Fig. 5. Definitions and Definition.

Those Definitions (Figure 4) will be used for giving additional statements in the following manner (Figure 5) in which two kinds of statements will be accepted, either the assignment of a value to a variable already defined or the result of an algebraic expression. In this paper, the notion of algebraic expression will not be defined since it is very common in all mathematical languages. In addition, in the Definitions, it is possible to state a relationship between variables considered as a sort of constraint.

Now that the Context is fully characterized, let's consider Conditions. Similarly, Conditions and Condition will be defined in Figure 6.

Regarding implications, as previously told, three modalities exist (Figure 7) in which BP means best practices. As usual, Support and Confidence are percentages for rules coming from data mining.

Now, regarding Consequents, the structure will be similar but recursive (Figure 8).

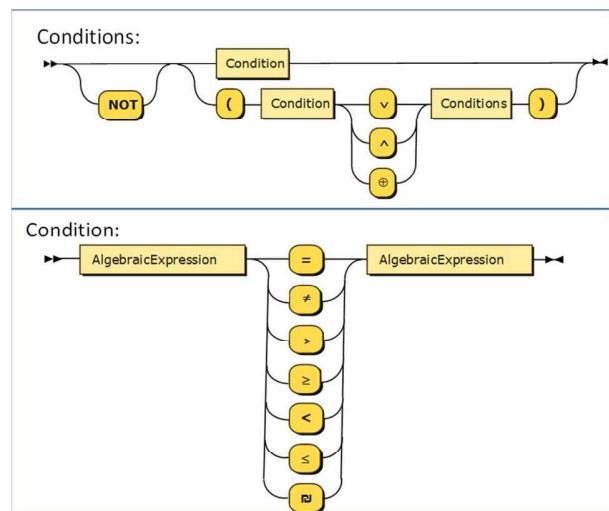


Fig. 6. Conditions and Condition.

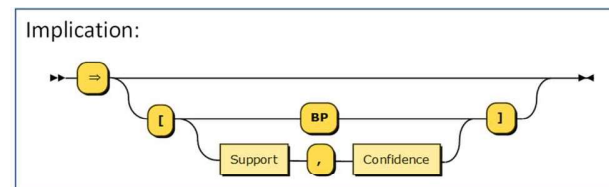


Fig. 7. Implication with three modalities.

In this grammar, the main concepts were presented. Some concepts such as “variable”, “Algebraic Expression”, etc. have their usual meaning in mathematics. Now, let us consider a few examples.

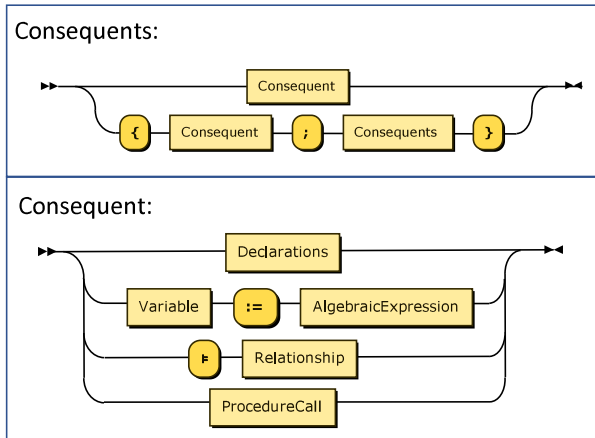


Fig. 8. Consequents and Consequent.

6. Examples

Here are given a few examples of rules written with the above-mentioned language. For that, we suppose the existence of a town named "Smart Town".

6.1 Creating a new relation, a new zone

For defining a new relationship, it can be easily done, for instance for *North* (see Rule 1). An equivalent could be written for *South*. This rule is valid anywhere in the *Earth*.

$\begin{aligned} &\forall p_1, p_2 \in \text{Earth}, \text{GeomType}(p_1) \equiv \text{Point}, \\ &\quad \text{GeomType}(p_2) \equiv \text{Point} \\ &\quad \vdots \\ &\quad \text{Latitude}(p_1) > \text{Latitude}(p_2) \\ &\quad \Rightarrow \\ &\quad \models \text{North}(p_1, p_2) \blacksquare \end{aligned}$	Rule 1
---	-----------

For defining *East* and *West*, the rule is a little more complex (Rule 2).

$\begin{aligned} &\forall p_1, p_2 \in \text{Earth}, \\ &\quad \text{GeomType}(p_1) \equiv \text{Point}, \text{GeomType}(p_2) \equiv \\ &\quad \quad \text{Point} \\ &\quad \quad \vdots \\ &\quad \quad (\text{Longitude}(p_1) > \text{Longitude}(p_2)) \\ &\quad \quad \wedge (\text{Longitude}(p_1) - \text{Longitude}(p_2) < 180^\circ) \\ &\quad \quad \Rightarrow \\ &\quad \quad \models \text{West}(p_1, p_2) \blacksquare \end{aligned}$	Rule 2
---	--------

Suppose we want to create a new relation when a road is crossing a river. We need to consider a road, a river and their intersection. If there is an intersection between the road and the river, there is a *cross* relationship between them, and it is commutative (Rule 3).

$\begin{aligned} &\forall Ro \in \text{Road}, \forall Ri \in \text{River} \\ &\quad \vdots \\ &\quad \text{Area}(\text{Intersection}(\text{Geom}(Ro), \text{Geom} \\ &\quad \quad (\text{Ri}))) \neq 0 \\ &\quad \Rightarrow \\ &\quad \{ \models \text{Cross}(Ro, Ri); \models \text{Cross}(Ri, Ro) \} \blacksquare \end{aligned}$	Rule 3
---	-----------

For the creation of a flood-risk area in Smart Town, let's suppose it to be defined within a 100 m buffer. Beware, the river can pass either through the town (*Overlap*) or be at the border (*Covers*). Here we must use a constructor, that is a procedure to create a novel geographic object (*CreateGO*). See Rule 4).

$\begin{aligned} &\exists T \in \text{Town}, \forall R \in \text{River}, \text{Topo}(T) \equiv \\ &\quad \quad \text{"Smart Town"} \\ &\quad \quad \vdots \\ &\quad \quad \text{Overlap}(R, T) \vee \text{Covers}(T, R) \\ &\quad \quad \Rightarrow \\ &\quad \quad \{ \text{CreateGO}(F); \text{Type}(F) := \text{"Floodplain"}; \\ &\quad \quad \quad \text{Geom}(F) := \text{Intersection}(\text{Geom}(T), \text{Buffer} \\ &\quad \quad \quad (R, 100)) \} \blacksquare \end{aligned}$	Rule 4
--	-----------

Suppose that a law decides that when a brownfield site is depolluted, it is possible to accept a new building (Rule 5).

$\begin{aligned} &\forall P \in \text{Parcel}, L \in \text{Plants}, B \in \\ &\quad \quad \text{Project. Building}, \\ &\quad \quad \models \text{Contains}(P, L), \models \text{Contains}(P, B) \\ &\quad \quad \quad \vdots \\ &\quad \quad \quad L.\text{Erased} \wedge P.\text{Depolluted} \\ &\quad \quad \quad \Rightarrow \\ &\quad \quad \quad \models B.\text{Authorized} \blacksquare \end{aligned}$	Rule 5
--	-----------

Suppose now that it was decided not to build in a marsh: this rule can be written as follows (Rule 6):

$\begin{aligned} &\exists C \in \text{County}, M \in \text{Marsh}, \forall B \in \text{Project}, \\ &\quad \quad \models \text{Contains}(C, M) \\ &\quad \quad \quad \vdots \\ &\quad \quad \quad \text{Contains}(M, B) \\ &\quad \quad \quad \Rightarrow \\ &\quad \quad \quad \text{Prohibit}(B) \blacksquare \end{aligned}$	Rule 6
--	--------

6.2 Again, the More, the Merrier

Suppose that "Smart Beach" is a sea resort in which land prices diminish when the distance to the sea increases. This economic rule (Rule 7) can be written as follows when considering two parcels with the same area:

$\begin{aligned} &\forall p_1, p_2 \in Parcel, \exists S \in Sea, C \in County, \\ &\quad \vDash Touches (Sea, C), \\ &\quad \vDash (Area (Geom (p_1)) \sqsupseteq Area (Geom (p_2))) \\ &\quad \vdots \\ &\quad Contains (C, p_1) \wedge Contains (C, p_2) \\ &\quad \wedge (distance (centroid (geom(p_1), S)) > \\ &\quad \quad distance (centroid (geom(p_2), S))) \\ &\quad \Rightarrow \\ &\quad \vDash p_1.Landprice < p_2.Landprice \blacksquare \end{aligned}$	Rule 7
---	--------

With the second solution (Rule 8), we get:

$\begin{aligned} &\forall p \in Parcel, S \in Sea, C \in County, \\ &\quad \vDash Touches (Sea, C) \\ &\quad \vdots \\ &\quad Contains (C, p) \wedge ++distance (centroid \\ &\quad \quad (Geom(p), S)) \\ &\quad \Rightarrow \\ &\quad \vDash -- p.Landprice \blacksquare \end{aligned}$	Rule 8
---	--------

Suppose that some economists have evaluated a function F for prices, provided that the area is more than 100 m² and the distance to the sea less than 10 km, the following can be written (Rule 9):

$\begin{aligned} &\forall p \in Parcel, \exists S \in Sea, C \in County, \\ &\quad \vDash Touches (Sea, C), \\ &\quad d \equiv (Centroid (Geom(p), S)) \\ &\quad \vdots \\ &\quad Contains (C, p) \\ &\quad \wedge Area (Geom (p)) < 100 \\ &\quad \wedge d < 10\ 000 \\ &\quad \Rightarrow \\ &\quad p.Landprice := F (Area (geom (p)), d) \blacksquare \end{aligned}$	Rule 9
---	--------

6.3 With fuzzy attributes

In the previous case, another solution can be based on fuzzy values: the previous rule can be transformed into "near the sea, prices are high", and "far from the sea, prices are low" (Rule 10).

$\begin{aligned} &\forall p \in Parcel, \exists S \in Sea, C \in County, \\ &\quad \vDash Touches (Sea, C), \\ &\quad d \equiv (Centroid (Geom(p), S)) \\ &\quad \vdots \\ &\quad Dist (S, p) = "near" \\ &\quad \Rightarrow \\ &\quad p.Landprice := "high" \blacksquare \end{aligned}$	Rule 10
--	---------

An additional rule can be written, by respectively replacing "near" and "high" with "far" and "low".

6.4 Best practice rules

Suppose it is a best practice to assign a pollution sensor to each lamppost located in "Churchill Road"; the example is given in Rule 11.

$\begin{aligned} &\exists T, \in Town, \exists R \in Road, \forall L \in Lamppost, \\ &\quad \forall S \in Pollution.Sensor, \\ &\quad Topo (T) \equiv "Smart Town", \\ &\quad Topo (R) \equiv "Churchill Road" \\ &\quad \vdots \\ &\quad Contains (Geom(T), Geom(R)) \wedge Contains \\ &\quad \quad (Geom (R), Geom (L)) \\ &\quad \Rightarrow [BP] \\ &\quad Assign (S, L) \blacksquare \end{aligned}$	Rule 11
--	---------

6.5 Associative spatial rules

When studying road incidents in the city of Helsinki, Finland, Karasova et al. [8] have shown by spatial data mining that many incidents occur near bars and restaurants. More exactly, around each incident they have designed a 50 m buffer and see whether there were incidents in those zones (Support 1.7% and confidence 40.0%).

The fact that the support is weak means overall that in their database, there are many other objects, whereas the confidence level means that 40 % of incidents happen in the vicinity of bars or restaurant. This associative rule can be written as follows in which $Terr$ is a territory (Rule 12):

$\begin{aligned} &\exists C \in City, \forall B \in Bar, \forall R \in Restaurant, \\ &\quad \forall I \in Incident, \\ &\quad \exists RiskyZone \in Terr, Topo (C) \equiv "Helsinki", \\ &\quad \vDash Contains (Geom (C), Geom (B)), \\ &\quad \vDash Contains (Geom (C), Geom (R)), \\ &\quad \vDash Contains (Geom (C), Geom (I)), \\ &\quad Geom (RiskyZone) \equiv \\ &\quad Union (Buffer (Centroid (Geom(B), 50), \\ &\quad \quad Buffer (Centroid (Geom(R), 50))) \\ &\quad \Rightarrow [1.7%, 40.0%]) \\ &\quad \vDash Contains (Geom (RiskyZone), Geom \\ &\quad \quad (I)) \blacksquare \end{aligned}$	Rule 12
--	---------

In their study in the city of Antwerp, Belgium, Zhou et al. [23] have a lot of co-location association rules within 600 m buffers.

For instance, they discovered such a rule between kindergartens, playgrounds, but support and confidence were not mentioned (let write α and β for those unknown values in the Rule 13).

$\begin{aligned} &\exists C \in City, \forall K \in Kindergarten, \exists P \\ &\quad \in Playground, \\ &\quad Topo(C) \equiv "Antwerp" \\ &\quad : \\ &\quad Contains(Geom(C), Geom(K)), \\ &\quad Contains(Geom(C), Geom(P)) \\ &\quad \Rightarrow [\alpha, \beta] \\ \models &Contains(Geom(Centroid(Geom(K), 600), \\ &\quad Geom(Centroid(P))) \blacksquare \end{aligned}$	<p>Rule 13</p>
--	--------------------

6.6 Back to previous rules

Let us now rewrite rules given in §3.

The example previously given from Malerba et al. [13] from an Italian census can be modeled by (Rule 14):

$\begin{aligned} &\exists P \in Earth, \forall X \in LargeTown, \forall Y, Z \in \\ &\quad Road, \\ &\quad Topo(P) \equiv "Italy" \\ &\quad : \\ &\quad Intersection(X, Y) \wedge Z \neq Y \\ &\quad \Rightarrow [45\%, 90\%] \\ \models &Intersection(X, Z) \blacksquare \end{aligned}$	<p>Rule 14</p>
--	--------------------

The first rule given by Assab et al. [18] (*Mines* – \exists_{3km}^3 *Faults* → *True*) can be translated into (Rule 15):

$\begin{aligned} &\exists SA \in Earth, \exists^3 F \in Fault, \\ &\quad Topo(SA) \equiv "South America" \\ &\quad : \\ &\quad Contains(Geom(SA), Geom(F)) \\ &\quad \Rightarrow \\ &\quad \{\exists M \in Mines; \\ \models &Contains(Geom(Buffer(Centroid(M), \\ &\quad 5000), Geom(F))) \blacksquare \end{aligned}$	<p>Rule 15</p>
--	--------------------

To finish this section, let's consider the ultimate rule designed by Varadharajulu et al. [22] (Rule 16):

$\begin{aligned} &\forall R \in Project.Road \\ &\quad : \\ &\quad R.Length(R) < 200, Type(R) = "Close" \\ &\quad \Rightarrow \\ \models &R.Allowed \blacksquare \end{aligned}$	<p>Rule 16</p>
---	--------------------

6.7 Located rules

Sometimes rules are located in small places, for example for urban planning zones. Figure 9 illustrates the case in which there are 3 zones in which 5 rules can be applied, those zones being delimited by the list of surrounding streets. Two solutions are possible:

- 1 – for each zone, give the list of applicable rules,
- 2 – for each rule, give the list of zones in which they apply.

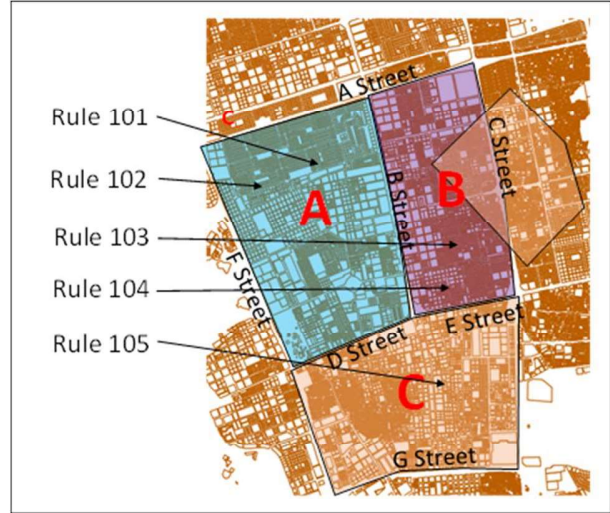


Fig. 9. Rules and zones.

With this rule-oriented language, let's detail take the option that, for each zone, will give the rules to be applied. For zone A, the following can be written (Rule 17).

$\begin{aligned} &\exists C \in City, \forall B \in Project, \exists ZoneA \in Terr, \\ &\quad Geom(ZoneA) \equiv SurroundedByStreet \\ &\quad (A_Street, B_Street, D_Street, F_Street) \\ &\quad : \\ &\quad Contains(Geom(ZoneA), Geom(B)) \\ &\quad \Rightarrow \\ \models &\{AppliedRule(101); AppliedRule(102)\} \blacksquare \end{aligned}$	<p>Rule 17</p>
---	--------------------

And similar rules can be written for ZoneB and ZoneC.

6.8 Superseded rules

As given in Figure 1, geographic rules can be superseded. But this expression has two meanings. Let us explain them.

The first means that a rule can be replaced by another rule or a set of new rules. For instance, in urban planning when a new plan is adopted. In this case, the previous rule is removed and a new one is enabled. However, geographic objects generated by means of the previous rule still continue to stay; in addition, suppose the previous rule imposes a limitation of 20 m for buildings and the new one only 15 m; it does not imply to demolish those 20m-high existing buildings. Let's call this, temporal superseding.

The second meaning is more complex: suppose that a

new rule is now valid in a small portion of a territory as illustrated in Figure 10. Two solutions are possible; either to state that the new rule (in Z_2) has a priority over the rule in Z_1 ; or to modify the initial rule by replacing $Geom(Z_1)$ by $Minus(Geom(Z_1), Geom(Z_2))$. Let's call this, spatial superseding.



Fig. 10. Example of spatial superseding.

7. Example in Urban Planning

Building permit delivery is a very complex task for local authorities. Indeed, a projected building must be compliant with two types of constraints, national laws and local rules. National laws which are valid everywhere in the country, generally consider structural aspects, such as electricity, insulation, access for handicapped people, etc. – issues which are outside the goal of this paper – whereas local rules consider the location of the building and its conformance to local planning rules, such as height, distance to neighboring parcels, floorspace ratios, etc. Figure 11 depicts the map with the extension of planning zones, and Table 1 gives some limit parameters like maximum height, maximum floorspace ratio and maximum footprint of the building.

TABLE I. PLANNING ZONES AND THEIR PARAMETERS

Zone ID	Max Height (in m)	Max Floorspace ratio	Max Footprint
Downtown	12	3	80 %
Suburban area	15	4	70 %
Rural area	12	0.5	30 %
Near airport (Bowtie)	8	2	50 %
Airstrip	0	0	0

7.1 Rules for zones

First let's suppose we have rules for dividing the territory and creating zones.

And so, for *ZoneB*, "Suburban area"; *ZoneC*, "Rural

area"; *ZoneD*, "Near airport" and *ZoneE*, "Airstrip", with the appropriate coordinates, respectively numbered Rule 19, Rule 20 and Rule 21. For a projected building supposedly located in the Suburban Area, the rule will be as follows to be accepted.

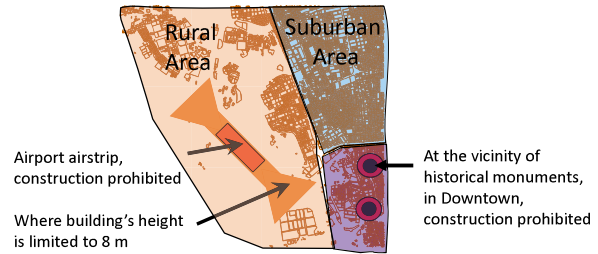


Fig. 11. Planning zones

$\exists C \in City, \exists PZoneA \in Project.Terr,$ $Topo(C) \equiv \text{"Smart Town"}$ $Geom(PZoneA) \equiv Poly(731, 128; 903, 133;$ $905, 341; 839, 346; 814, 349)$ \vdots $Approved(PZone)$ \Rightarrow $AffectName(PZoneA, \text{"Downtown"}) \blacksquare$	Rule 18
---	----------------

In this toy-example, we will consider a projected building presented to "Smart Town" local authority, the urban planning department of which has decided to cut the town into three zones, namely downtown, suburban area and rural area.

7.2 Rules regarding parcels and construction

To simplify, we will not consider parcels in which construction is not possible, and parcels for which municipalities may trigger some pre-emptive rights (for instance for schools, parks, etc.).

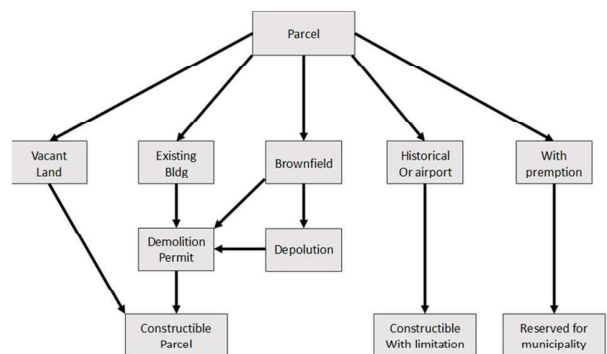


Fig. 12. Different states for a parcel regarding planning.

When there is a pre-existing building, a demolition permit must be granted and when it was a plant

(brownfield), depollution is need (Figure 12). In addition, at the vicinity of a historical monument or of an airport, special limitations apply.

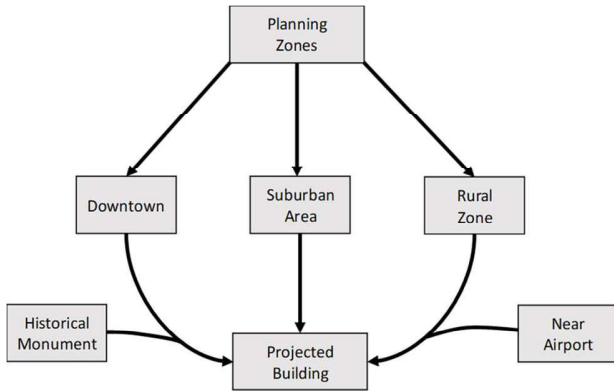


Fig. 13. Any projected building must be compliant with planning rules.

7.3 Projected Buildings and Urban Planning Rules

Regarding planning zones, each has its own regulations. For example, campsites are not allowed in downtown. To simplify, three zones are defined, downtown, suburban area and rural area. However, historical monuments are located in downtown, and it is forbidden to construct new buildings round them (for instance, within 200 meters). In the rural area, there is an airport for which some limitations exist (Figure 11). See Rule 22.

$\begin{aligned} &\exists C \in City, \exists ZoneB \in Terr, \forall B \in Project.Building, \forall P \in Parcel, \\ &Topo(C) \equiv \text{"Smart Town"}, \\ &Topo(ZoneB) \equiv \text{"Suburban Area"} \\ &\models Contains(Geom(C), Geom(ZoneB)), \\ &\models Contains(Geom(ZoneB), Geom(P)), \\ &\models Contains(Geom(P), Geom(B)) \\ &\vdots \\ &B.Height \leq 15 \\ &\wedge Area(Union(Geom(Floors)))/Area(Geom(P)) \leq 4 \\ &\wedge Area(B)/Area(Geom(P)) \leq 0.70 \\ &\Rightarrow \\ &\models B.ZoneB_Approved \blacksquare \end{aligned}$	Rule 22
---	------------

But for Downtown and Rural Area, the situation is more complex. For Downtown, we have to take historical monuments into account, for instance within a 200 m buffer around its centroid. See Rule 23.

For the Rural Area, taking the airport into consideration, we need to consider three areas, the area outside the airport, the airstrip in which any building is forbidden, and finally in the “bowtie” with additional limits (Rule 24).

Technically speaking, we have to use an “exclusive or”, noted \oplus between those three possibilities.

$\begin{aligned} &\exists C \in City, \exists ZoneA, ConservA \in Terr, \forall B \in Project.Building, \\ &\forall P \in Parcel, \forall M \in Monuments, \\ &Topo(C) \equiv \text{"Smart Town"}, \\ &Topo(ZoneA) \equiv \text{"Downtown"}, \\ &Geom(ConservA) \equiv Union(Buffer(Centroid(Geom(M), 200))), \\ &\models Contains(Geom(C), Geom(ZoneB)), \\ &\models Contains(Minus(Geom(ZoneA), Geom(ConservA)), Geom(P)), \\ &\models Contains(Geom(P), Geom(B)) \\ &\vdots \\ &B.Height \leq 12 \\ &\wedge Area(Union(Geom(Floors)))/Area(Geom(P)) \leq 3 \\ &\wedge Area(B)/Area(Geom(P)) \leq 0.80 \\ &\Rightarrow \\ &\models B.ZoneA_Approved \blacksquare \end{aligned}$	Rule 23
--	------------

$\begin{aligned} &\exists C \in City, \exists ZoneC, Bowtie, Airstrip \in Terr, \\ &\forall B \in Project.Building, \\ &\forall P \in Parcel, \\ &Topo(C) \equiv \text{"Smart Town"}, \\ &Topo(ZoneA) \equiv \text{"Rural Area"}, \\ &Geom(Bowtie) = Polyg(640, 243; 657, 290; \\ &748, 387; 796, 405; 743, 459; 729, 406; 636, \\ &316; 580, 297), \\ &Geom(Airstrip) = Polyg(670, 311; 724, 365; \\ &707, 386; 650, 330), \\ &\models Contains(Geom(C), Geom(ZoneC)), \\ &\models Contains(Geom(P), Geom(B)) \\ &\vdots \\ &(Contains(Minus(Geom(ZoneC), Geom(Bowtie)), Geom(B)), \\ &\wedge B.Height \leq 12 \wedge Area(Union(Geom(Floors)))/Area(Geom(P)) \leq 0.5 \\ &\wedge Area(B)/Area(Geom(P)) \leq 0.30) \\ &\oplus Disjoint(Geom(Airstrip), Geom(B)) \\ &\oplus (Contains(Geom(Bowtie)), Geom(B)) \\ &\wedge B.Height \leq 8 \wedge Area(Union(Geom(B.Floor)))/Area(Geom(P)) \leq 0.5 \\ &\wedge Area(B)/Area(Geom(P)) \leq 0.30) \\ &\Rightarrow \\ &\models B.ZoneC_Approved \blacksquare \end{aligned}$	Rule 24
---	------------

Now, to be fully approved by the local administration, whatever its location, a projected building to be approved must follow the subsequent rule; here again, exclusive-ors (\oplus) must be used (Rule 25).

$\begin{aligned} & \exists C \in \text{City}, \forall B \in \text{Project. Building}, \\ & \text{Topo}(C) \equiv \text{"Smart Town"}, \\ & \models \text{Contains}(\text{Geom}(C), \text{Geom}(B)) \\ & \quad \vdots \\ & (\text{Contains}(\text{Geom}(\text{ZoneA}), \text{Geom}(B)) \wedge \\ & \quad \text{B.ZoneA_Approved}) \\ \oplus & (\text{Contains}(\text{Geom}(\text{ZoneB}), \text{Geom}(B)) \wedge \\ & \quad \text{B.ZoneB_Approved}) \\ \oplus & (\text{Contains}(\text{Geom}(\text{ZoneC}), \text{Geom}(B)) \wedge \\ & \quad \text{B.ZoneC_Approved}) \\ & \Rightarrow \\ & \models \text{B.FullyApproved} \blacksquare \end{aligned}$	<p>Rule 25</p>
--	--------------------

Of course, those rules can be enriched to take depollution for possible brownfields into account as exemplified in Rule 5, flood restriction, etc. Moreover, additional rejection rules can be written.

8. Conclusions

The goal of this paper was to present the first version of a mathematical language for geographic static rule modeling, independent from any computer language and able to integrate all semantics. Thus, this language is based on several mathematical domains such as logic, set theory, computational geometry, topology, etc.

At the moment, only several hundreds of 2D static rules are modeled with an interesting expressive power. For the complete modeling of any geographic rule, research must be carried out in several directions:

- integration of 3D issues, especially for terrain modeling and engineering networks;
- integration of temporal issues; this will lead to dynamic geospatial rules; remember that time semantics are very complex in geography, ranging from fuzzy billions of years in geology until seconds for real time sensors;
- integration of rules deriving from continuous fields, especially for dealing with meteorology, pollution, etc. and other aspects in physical geography;
- integration of additional clauses to extend its expressive power, overall to deal with networks whatsoever, electricity, sewerage, bus lines, etc.;
- looking for more issues in order to enrich semantics, especially for the automatic adaptation to special contexts; for instance, how to adapt a rule such as "when planning a metro, move underground engineering networks" to various street configurations;
- transformation of this mathematical language into a computer language;

- study of metadata relative to geographic rules (origin, modality, etc.);
- design of an inference engine to reason with those rules;
- defining the organization of rules together for their access mechanisms taking temporal and spatial superseding mechanisms.

To conclude this paper, let me thanks the anonymous referees for their work and especially their fruitful comments.

References

- [1] Boley H. (2006) "The RuleML Family of Web Rule Languages". International Workshop on Principles and Practice of Semantic Web Reasoning. Springer, Berlin, Heidelberg, pp. 1-17. Can be downloaded from <http://www.cs.unb.ca/~boley/papers/ruleml-family.pdf>
- [2] Boley, H., Paschke, A., Shafiq, O. (2010) "RuleML 1.0: The Overarching Specification of Web Rules". In Semantic Web Rules: International Symposium, RuleML 2010, Washington, DC, USA, October 21-23, 2010, Proceedings (Vol. 6403, p. 162). Springer Science & Business Media.
- [3] Dietz J.L.G. (2008) "On the Nature of Business Rules". In Advances in Enterprise Engineering, Springer Verlag Lecture Notes in Business Information Processing (10) pp. 1-15.
- [4] Egenhofer M. & Franzosa R.D. (1991) "Point-set topological spatial relations", International Journal of GIS, vol.5, no.2, pp. 161-174.
- [5] Espinasse B. (2017). "Introduction to Semantic Web Rule Language -SWRL". Can be downloaded from <http://www.lsis.org/espinasseb/Supports/ONTOWS/SWRL.pdf>
- [6] Graham I. (2006) "Business Rules Management and Service Oriented Architecture: A Pattern Language". London, John Wiley.
- [7] Kainz W. (2002) "Fuzzy Logic and GIS". Can be downloaded from https://homepage.univie.ac.at/Wolfgang.Kainz/Lehrveranstaltung/en/ESRI_Fuzzy_Logic/File_2_Kainz_Text.pdf
- [8] Karasova V., Krisp J.-M., Virrantaus K. (2005) "Application of Spatial Association Rules for Improvement of a Risk Model for Fire and Rescue Services". Proceedings on the 10th Scandinavian Research Conference on Geographical Information Science (ScanGIS), Stockholm, Sweden, pp. 183-193.
- [9] Laurini, R. (2014) "A Conceptual Framework for Geographic Knowledge Engineering", Journal of Visual Languages and Computing (2014), Volume 25, pp. 2-19.
- [10] Laurini R., Servigne S., Favetta F. (2016) "An Introduction to Geographic Rule Semantics". In Proceedings of the 22nd International Conference on Distributed Multimedia Systems (DMS 2016), Salerno, Italy, November 25-26, 2016. Published by Knowledge Systems Institute, ISBN: 1-891706-40-3, pp. 91-97.
- [11] Laurini, R. (2017) "Geographic Knowledge Infrastructure for Territorial Intelligence and Smart Cities". ISTE-Wiley. 250 p.
- [12] Levenshtein, V. I. (1966). "Binary codes capable of correcting deletions, insertions, and reversals". In Soviet Physics, doklady, volume 10, pp. 707-710.
- [13] Malerba D., Esposito F., Lisi, F., Appice A. (2003). "Mining Spatial Association Rules in Census Data. Research in Official Statistics. 5. Can be downloaded from https://www.researchgate.net/publication/2839474_Mining_Spatial_Association_Rules_in_Census_Data/download.

- [14] Morgan T. (2008) "Business Rules and Information Systems: Aligning IT with Business Goals". Addison-Wesley.
- [15] Pittl B., Fill H.-G. (2018) "A Visual Modeling Approach for the Semantic Web Rule Language". *Semantic Web Journal*. September 2018. Can be downloaded from <http://www.semantic-web-journal.net/system/files/swj2023.pdf>
- [16] Ross R. G. (2011) "More on the If-Then Format for Expressing Business Rules: Questions and Answers", *Business Rules Journal*, Vol. 12, No. 4 (Apr. 2011), URL: <http://www.BRCommun2002ity.com/a2011/b588.html>.
- [17] Rademacher G. (2019) "Railroad Diagram Generator". <https://bottlecaps.de/rr/ui>
- [18] Salleb-Aouissi A., Vrain C. & Cassard D. (2015) "Learning Characteristic Rules in Geographic Information Systems". N. Bassiliades, G. Gottlob, F. Sadri, A. Paschke, D. Roman (Eds.). *Rule Technologies: Foundations, Tools, and Applications*, 9th International Symposium, (RuleML 2015), Aug 2015, Berlin, Germany. Springer, 9202, 2015, *Lecture Notes in Computer Science*.
- [19] Smith B., Varzi A., (2000), "Fiat and Bona Fide Boundaries". *Philosophy and Phenomenological Research* Vol. LX, No. 2, March 2000, pp. 401-420.
- [20] Shekhar S., Huang Y., (2001) "Discovering Spatial Co-location Patterns: A Summary of Results", in *Proc. of the 7th Int'l Symposium on Spatial and Temporal Databases*, 2001, pp. 236–256.
- [21] Shoorcheh M. (2018) "On the spatiality of geographic knowledge". *Asian Geographer*. 36:1, pp. 63-80, DOI: 10.1080/10225706.2018.1463854.
- [22] Varadharajulu P., West G., Mcmeekin D., Moncrieff S. & Arnold L. (2016), "Automating Government Spatial Transactions". In *Proceedings of the 2nd International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM)*, held in April 2016, Rome, Italy, edited by Jorge Gustavo Rocha and Cédric Grueau, published by Scitepress, ISBN: 978-989-758-188-5, pp. 157-167.
- [23] Zhou C., Xiao W.D, Tang D.Q. (2016) "Mining Co-Location Patterns from Spatial Data". *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume III-2, 2016 XXIII ISPRS Congress, 12–19 July 2016, Prague, Czech Republic, pp. 85-90.

J