

Spatial Analysis to Generate a Conceptualization of Geometric Properties

Karina Verastegui, Marco Moreno-Ibarra and Miguel Torres
Geoprocessing Laboratory-CIC- National Polytechnic Institute, Mexico City, Mexico
verastegui@sagitario.cic.ipn.mx,
{marcomoreno, mtorres}@cic.ipn.mx
<http://geo.cic.ipn.mx>

Abstract. We propose an approach oriented to generate an explicit representation of the geometric characteristics of geospatial data. Our methodology is based on a conceptualization of a geospatial domain. The concepts are extracted automatically by analyzing the properties of geospatial objects. Also the concepts are stored in the spatial database to support subsequent processing. The conceptualization of geometric properties is based on measurements of geospatial objects. The results of this measurement are classified to obtain representative clusters of values in order to describe these properties. The values are used to define which concept better represents the properties of each object. We consider different geometric characteristics in a particular case study, such as rivers. Then, the characteristics that we have considered are the following: sinuosity, length, slope, and course. By using this method, we attempt to catch the semantics that contains geospatial data in the geometric context.

1 Introduction

This research has as main purpose to design a Geographic Information System (GIS) oriented to explicitly represent the geometry of geospatial data. The explicit representation is based on the conceptualization of the geometric properties. This is represented by means of concepts that describe different geometry. To quantify the characteristics, different measurements are computed to make a conceptualization.

In this paper, we present a conceptualization oriented to generate ontological information and thus structure or homogenize the databases. We are looking for a format of data representation, which does not depend on scale, format, references, etc. According to [2] a body of formally represented knowledge is based on a *conceptualization*: the objects, concepts, and other entities that exist in some area of interest and the relationships that hold among them. A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose [3].

Ontologies encode semantic relations between concepts and hence facilitate the detection of associations between related terms. In modern information systems based on ontologies, one seeks canonical descriptions of knowledge domains and associated classificatory theories.

Traditionally, geographic data models usually explicitly represent a set of basic objects, their geometry and their properties. But much of the geographic world's semantics appears in the relations linking objects [12] [8]. Nevertheless, most of these rela-

tions are not explicitly represented in data models describing geographic databases. Usually, these relations only implicitly appear when one is looking at a *display* of a geographic database [7]. Geographic Information Systems (GIS) handle geo-spatial data at different levels of details. Due to page limit our literature overviews stopped at this point.

In this work, we generate a conceptualization of geometrical properties of geospatial objects and explicitly represent them into spatial databases. The paper is organized as follows: Section 2 describes the measurements to quantify the geometric characteristics. Additionally, Section 3 points out the conceptualization method. Some preliminary results are shown in Section 4, and Section 5 sketches out our conclusion and future work.

2 Conceptualization of geometrical characteristics

This research has as main purpose to design a Geographic Information System (GIS) oriented to explicitly represent the geometry the geospatial data. The explicit representation is based on the conceptualization of the geometrical properties (See Fig.1). The conceptualization of geometrical properties is based on measurements of geospatial objects. The values obtained by the measurements are classified to obtain representative clusters of values in order to describe these properties. The values are used to define which concept better represents the properties of each object.

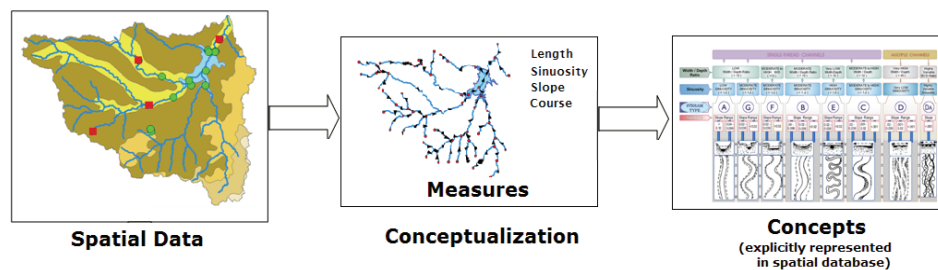


Fig. 1. Process to generate concepts explicitly represented

2.1 Measures

The necessity to measure is evident in most of the technical or scientific activities. Nevertheless, it does not only interest to count on measurements, but also to know if these measurements are valid. The validity of the measurement in any technical or scientific discipline is based on the respect to the principles of the general theory of the measurement (in particular, we will lean in the call representational theory of the measurement). The foundation of the representational theory consists of which all measurement must assure a suitable representation of the real attribute measured by

means of the symbols or assigned numbers. Additionally, the relationships should be preserved among this entity.

In order to establish measurements, we must divide our observation of the real world or dominion. We must identify which are the organizations that we want to measure (e.g., river) and to define what attribute we wished to characterize (e.g., the sinuosity). The measurement assigns a value to each organization to characterize its attribute (e.g., the river S3 has a sinuosity of 2.75). The most important is that the measurement that we establish is not inconsistent with the relations observed in the real world.

It is important to consider that not always the ideas on the attributes are so clear. We can begin by simple subjective valuations (e.g., to use questionnaires where they are classified or orders the opinions of the experts on an attribute), that do not constitute measurements from the point of view of the theory of the representation, but that they can be analyzed to improve the understanding on the real world. It is possible that after accumulating data of this type, it is necessary to be obtained to define a formal measurement. An allocation that settles down between real world and values of measurement usually is denominated measurement scale.

To study with detail the geometry of the spatial objects different simple algorithms should be developed, in order to reflect in some aspects the basic characteristics of the geometry. For example, in the generalization of a hydrological network some such geometric problems exist as lines that you angular and they lose detail, as well as nodes that are displaced as is shown in Fig. 2. In the generalization of a map, it takes to those contradictions exist between the demands of geometric accuracy and the geometric coincidence. The geometric exactitude of a map presupposes that each object of the terrestrial surface is exactly represented in the map in the same place with its true plane contours and dimensions, conserving the positions of all the objects and the distances that overcome them, correctly of agreement with the scale of the map.

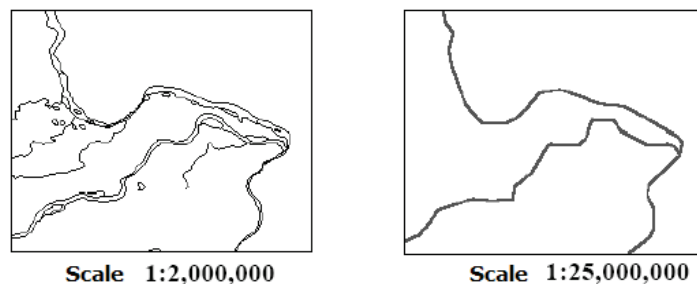


Fig. 2. Generalization of hydrological network

This reason is important to analyze, the change that occurs in the geometric form in to the objects. A form to make this process is by means of the measurement. In other works, this is the base of value the geometric characteristics in the generalization. In the Technical Annex of the project AGENT a classification of geometric measurements is shown and it is focused on the generalization. With this classifica-

tion will be carried out measurements of the geographical objects with the purpose of obtaining a conceptualization, guided to evaluate the result of the generalization.

In conclusion the measurements are numeric values assigned to an observation that reflects a magnitude, quantity or a characteristic. For this reason, we can make measurements according to their geometric form.

3 Methodology

The explicit representation is based on the conceptualization of the geometrical properties. This method does not depend on the scale as the traditional GIS approaches. In the following subsections, we outline our methodology.

3.1 Methodology of geometrical conceptualization

In our model, the geometry of geospatial objects will be described by means of concepts that represent geometrical properties, such as size, sinuosity / complexity, elongation / eccentricity, compactness, as well as other important aspects. We consider three types of geospatial data; point-like objects (i.e. well or tree), linear objects (i.e. a road or river) and the area objects (i.e. state boundary or lake).

The conceptualization process consists of four steps: (1) to obtain the measurements, we use algorithms that evaluate geometric characteristics. Fig. 3 shows a classification of measurements. Shape describes the geometric representation of spatial objects according to [11].

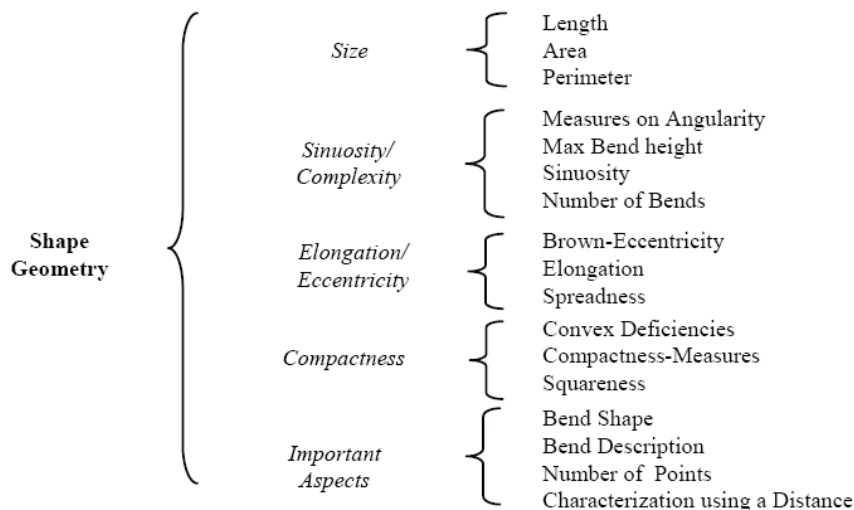
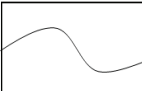
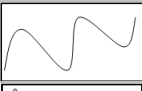
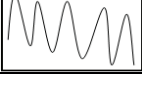


Fig. 3. Classification of Measures

(2) The obtained values by the measurements are classified to generate representative clusters of values in order to describe the geometric characteristics in a qualitative way (3) each cluster will be used to assign a concept according to the classification of intervals of the measurements (see Table 1). Fig. 4 depicts a workflow diagram of the conceptualization process.

Table 1. Mapping between the obtained values of a measurement and the concept

Spatial representation		Range (measure)	Concept
		1 - 1.4	Straight
		1.5 – 2.0	Little sinuous
		2.1 – 2.4	Half sinuous
		2.5 – 3.5	Very sinuous

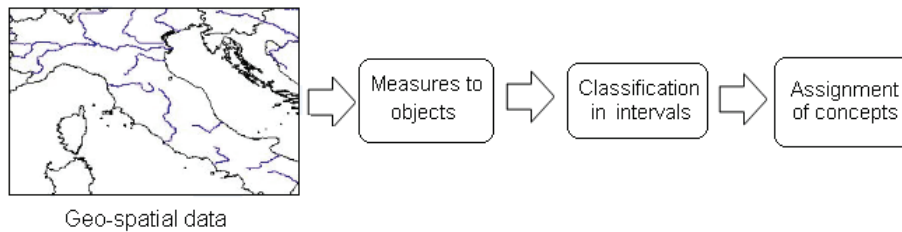


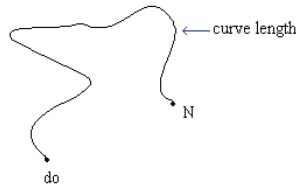
Fig. 4. Workflow diagram to obtain the conceptualization

3.2 Methodology of geometrical conceptualization of rivers

In this investigation was carried out the conceptualization of rivers by means of measurements (e.g. length, sinuosity, heights, trajectory and address of rivers), the values obtained by the measurements are classified in intervals, assignment of concepts according to the classification of intervals of the measurements.

3.2.1 Compute of length measurements

The length measurement (L), it is obtained by the equation 1, which is computed by the shortest distance between a point and a straight line is perpendicular one to this straight line that goes to that point. In other words, if we think of vectors instead of lines, it would be the vector that goes to the point of measurement of the distance that maintains a product to climb = 0.



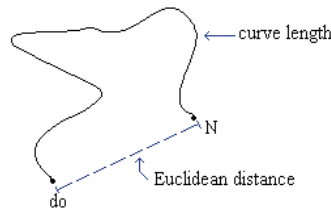
$$d_{(P, L)} = \frac{(y_0 - y_1)x + (x_1 - x_0)y + (x_0y_1 - x_1y_0)}{\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}}$$

Fig. 5. Compute of length measure

3.2.2 Compute of sinuosity measurement

In the case of *sinuosity measurement* (S), it is obtained by the equation 2, which is computed by dividing curve length with Euclidean distance from the initial node d_0 to the terminal node N ($P(d_0, N)$) (See Fig. 6a).

For example, S for different lines (L_0, L_1) is $S(L_0) = 1.0034$ and $S(L_1) = 3.3905$. It means that this property will be represented by the concepts *straight* and *very sinuosity* respectively (See Fig. 6b).



$$\text{Sinuosity} = \frac{\text{Curve length}}{P(d_0, N)}$$

a)

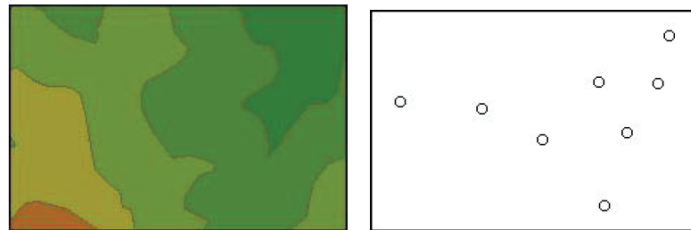
Classification of S	Concept
1.0 – 1.4	Straight
1.5 – 2.0	little sinuous
2.1 – 2.4	half sinuous
2.5 – 3.5	very sinuous

b)

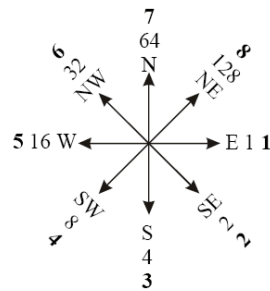
Fig. 6. Compute of sinuosity measurement

3.2.3 Compute of heights, trajectory and address of rivers Measure

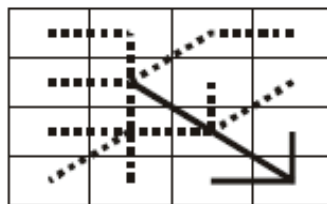
The slope of the land is computed by starting from the gradient of heights among neighboring cells (Fig 7a). The line of flows describes the address of the glides of water in the surface (Fig 7b), following the trajectory of more slope. To code the address of the glide, a code is assigned to each cell, among a total of 8 different values; each value represents the orientation of the cell, following the sketch: E=1, SE=2, S=4, SW=8, W=16, NW=32, N=64, NE=128 (Fig. 7c).



a) Compute of heights



c) Direction of flood



b) Compute of trajectory

Fig. 7. Compute of heights, trajectory and address of rivers

3.2.4 Classification in subsystems

To carry out a classification in everything the one laughed it is necessary to make a grouping for subsystems, a subsystem is the group of rivers defined by a watershed, each one of these areas will be denominated by an unique badge, for what we will have several subsystems in a group of rivers.

Horton in [4] and [5] introduced a concept of classification of streams that allows assigning whole values to streams in hydrological networks that determine their relative importance in a hierarchy of tributary bigger and smaller. An improved version of this concept was introduced later in [10]. The network is integrated by a main channel and a series of tributary whose ramifications extend from the highest parts toward the lowest parts where the glides converge [1] [9]. Fig. 8 shown a group the rivers in subsystems, with a classification of streams applying the hierarchical scheme of Horton and Strahler (HS).

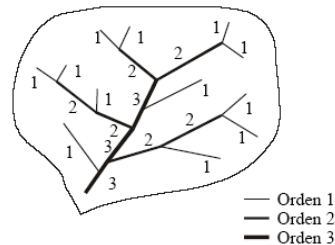


Fig.8. Classification of streams with the outline of Horton and Strahler

Once the rivers are classified in subsystems, the measurements are computed, corresponding with their sinuosity, length, order of the flowing (hierarchical class), and the range of heights.

4 Experimental results

To obtain the measurements, we have been developed in language AML (Arc Macro Language), a set of programs to obtain the geometric characteristics. In Fig. 9, a fragment of drainage network is depicted. It is composed of lineal elements, these lineal elements possess the characteristics of any other drainage network.

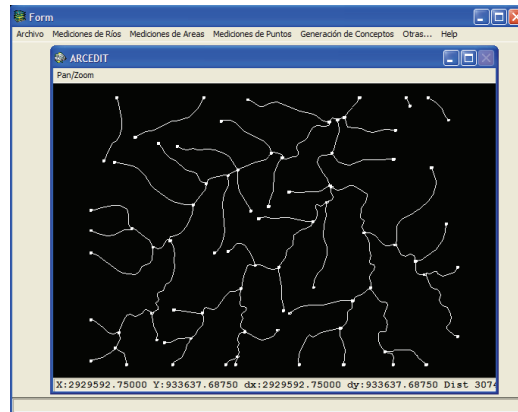


Fig. 9. Proposed drainage network

Fig. 10 (a) depicts the computation of *sinuosity measurement* for the Rivers. In addition, Fig. 10 (b) shows the compute of length measurement for the rivers.

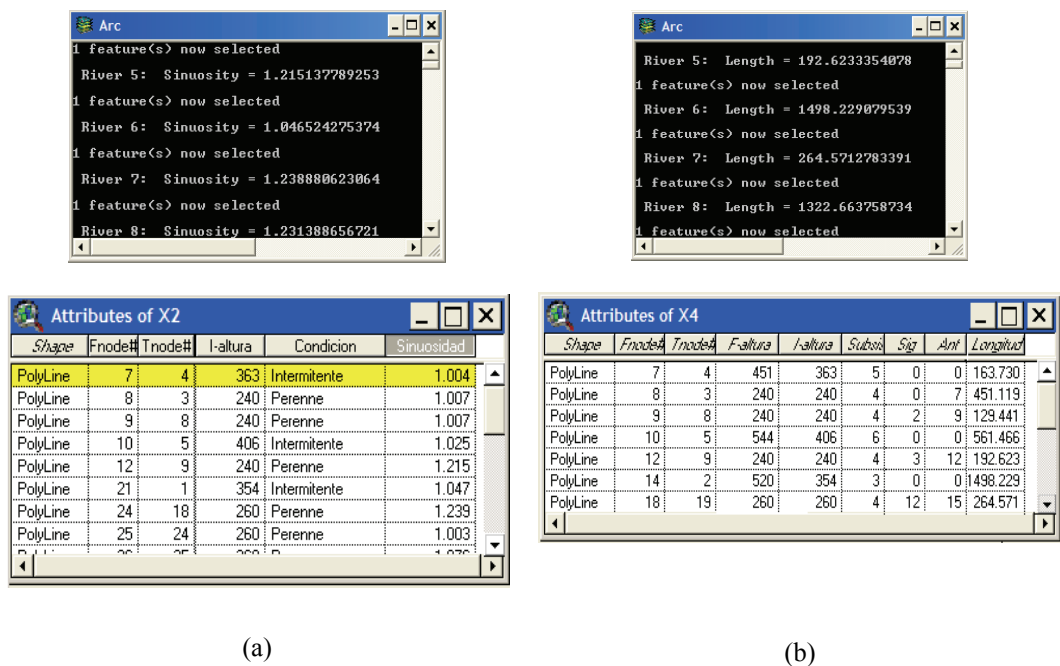


Fig. 10. (a) Sinuosity measurement, (b) Length measurement

The process of altitudes assignment, involves the use of the layer that contains the altitudes of that area, which is required to assign the heights to the nodes and later on.

This layer is used to extract the information on the altitudes. Fig. 11 depicts the computation of *heights* for the rivers.

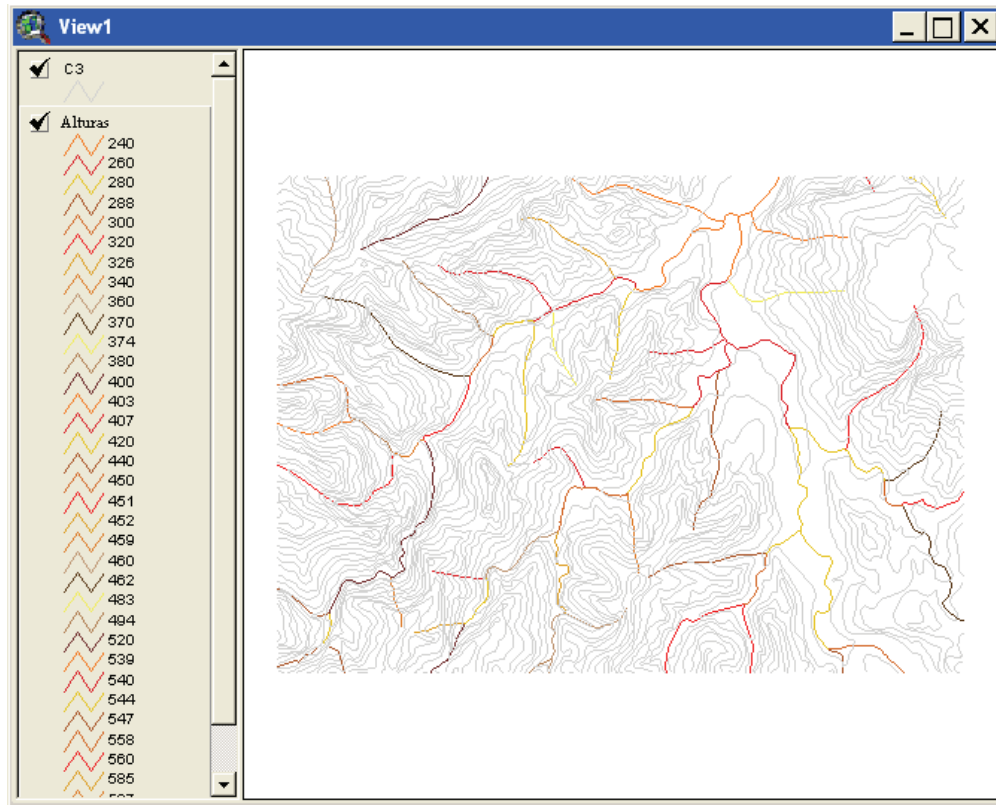


Fig. 11. Compute of heights of rivers

Fig. 12 depicts the classification of *subsystems* for the Rivers. This classification provides the measurements for subsystems and to make comparison of results later on.

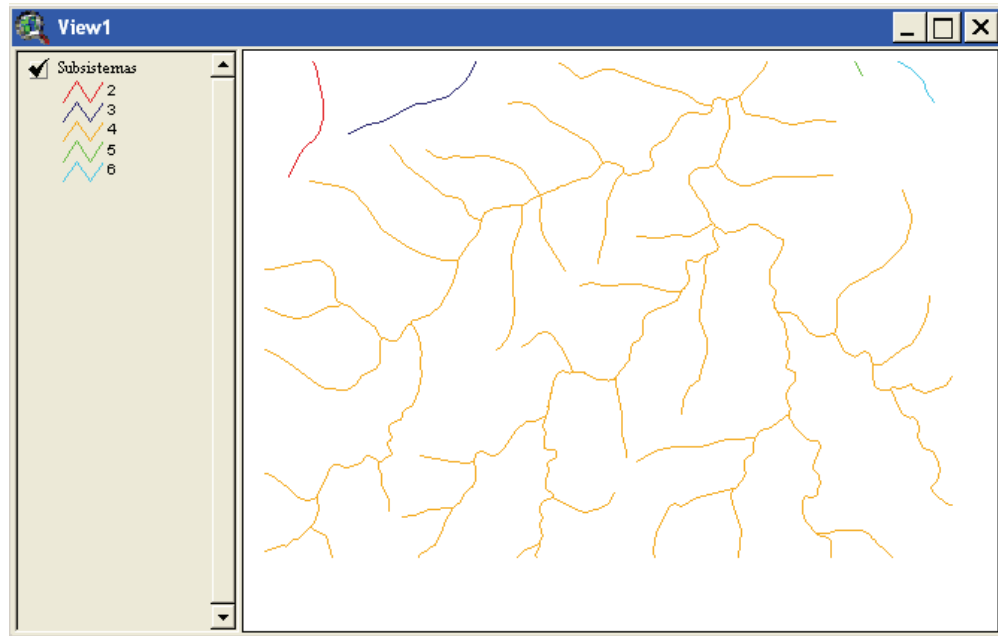


Fig. 12. Classification in subsystems

The following step is the *conceptualization*. The conceptualization process will be performed by a mapping between the range of values and the concept. In this case, the range of values represents a concept. The representative intervals are defined in the classification (for example, in the case of sinuosity a line can be *little*, *half*, *very sinuous*). The values and concepts are different according to the case study. In other words, each thematic should be represented by different intervals or concepts. Fig. 13 shows the mapping between the value of a measurement and the concept. This is defined by means of intervals (ranges).

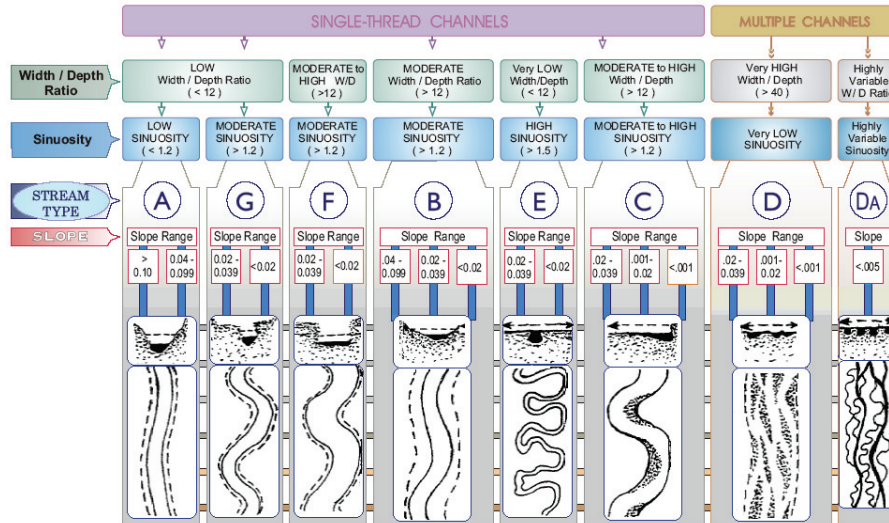


Fig. 13. Mapping between the obtained values of a measurement and the concept

It is important to mention that the generated concepts are not *ambiguous*, because they only can be applied in a specific context (geometric shape), i.e. it is different the sinuosity of highways that the sinuosity of rivers or elevation contour lines.

5 Conclusions and Future Work

In this work an approach to incorporate *semantic content* into Geographic Information Systems has been presented. The semantic data content is expressed by concepts. These concepts are generated analyzing the datasets in a specific geographic domain, and represented by concepts. The concepts represent the interpretation, and the meaning of geospatial objects. Our method is based on explicit representation of geometrical properties of geographic data by means of concepts. By using this approach, we attempt to catch or dimension *semantic contents*, which implicitly contain the spatial data and do not depend on the scale. In addition, it is possible to hierarchize the objects according to geometric characteristics.

In the future work we will measure the semantic similarity among different geospatial datasets. In our opinion, a domain conceptualization is useful to build ontologies, which represent (globally) the context of that domain, while the vocabulary of concepts and its relations describe the semantics (locally). Ontologies are very useful since they add a semantic component (the relations between different concepts), which normally is not considered in traditional GIS approaches.

Acknowledgements

The authors of this paper wish to thank the IPN, CIC, SIP and CONACYT for their support.

References

1. Black, P.E.: *Watershed hydrology*. Michigan, Ann Arbor Press, pp. 449, (1996).
2. Genesereth, M. R. and Nilsson, N.: *Logical Foundation of Artificial Intelligence*. Morgan Kaufmann, Los Altos, California, (1987).
3. Gruber T.: *Toward principles for the design of ontologies used for knowledge sharing*, International Journal of Human-Computer Studies, Volume 43 , Issue 5-6 Nov./Dec. 1995 table of contents, Special issue: the role of formal ontology in the information technology, Pages: 907 - 928, Year of Publication, ISSN:1071-5819, (1995).
4. Horton, R.E.: *Drainage basin characteristics*, American Geophysical Union Transactions, pp. 350-361, (1932).
5. Horton, R.E.: *Erosional development of streams and their drainage basins: hydrophysical approach to quantitative morphology*. Geol. Soc. America Bull., 56, 275-370, (1945).
6. Llamas, J.: *Hidrología general, principios y aplicaciones*, Universidad Autónoma del Estado de México, pp.626, (1989).
7. Mustière S.: What is spatial context in cartographic generalisation? *Symposium on Geospatial Theory, Processing and Applications* (2002).
8. Papadias D. and Theodoris Y.: Spatial relations, minimum bounding rectangles, and spatial data structures, *International Journal of Geographical Information Science*, vol. 11, n.2, pp.111-138 (1997).
9. Sanjuane, M.S. y Villanueva, R.J.B.: *Teoría y métodos en geografía física*, Editorial Síntesis, Madrid, pp 303, (1996).
10. Strahler, A.N.: *Quantitative analysis of watershed geomorphology*. Am. Geophys. Union Trans., 38(6), p. 913-920, (1957).
11. UNI-ZH.: AGENT Technical Annex [IGN, 96], Dept. of Geography, University of Zürich (1999).
12. Worboys M.F.: Metrics and topologies for geographic spatial. In *Advances in GIS research II : proc. of 7th International Symposium on Spatial Data Handling*, Kraak and Molenaar (eds), Taylor and Francis, pp.365-375, (1996).