We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Fuzzy Spatial Data Warehouse: A Multidimensional Model

Pérez David, Somodevilla María J. and Pineda Ivo H.
Facultad de Ciencias de la Computación, BUAP,
Mexico

1. Introduction

A data warehouse is defined as an integration of a subject-oriented, a time-variant, and a non-volatile data. Based on the definition, we organize the data warehouse by domains (spatial areas) and in thematic categories (types of features) [1][2].

The loading and maintenance processes are one of the tasks that more effort and demand require. The ETL process needs a temporary storing place for the recovered data from its sources (transactional databases). In our study case, we have that the agency in charge of the study of Popocatepetl's behavior obtains the information from different sources as it's shown in Figure 1.
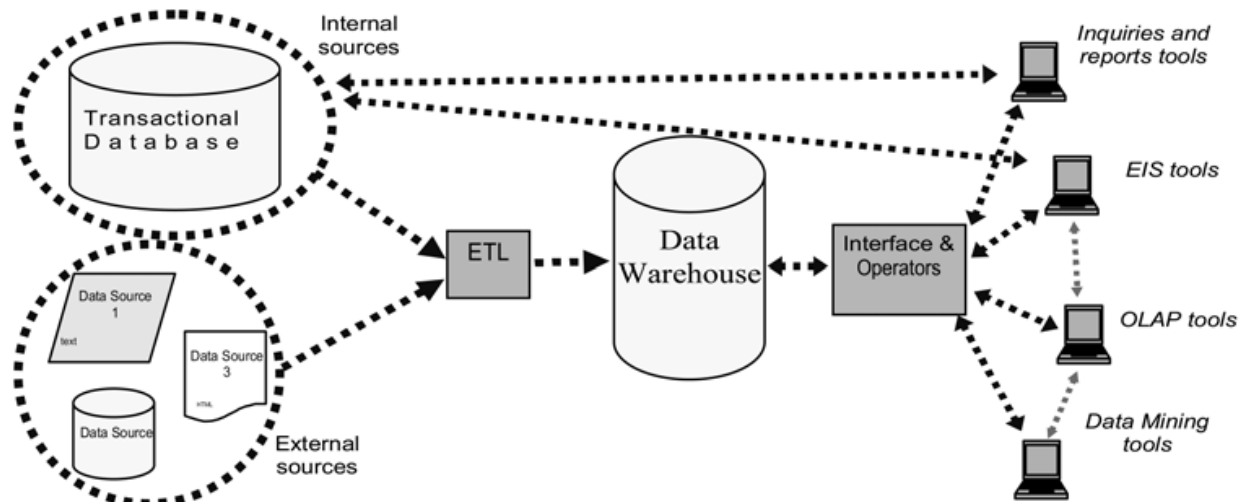

Fig. 1. Data Warehouse's Arquitecture

We present an object model for defining vague regions which rests on "traditional" (that is, exact) modeling techniques. This modeling strategy simultaneously expresses the authors' opinion that it is unnecessary to begin from scratch when modeling vague spatial objects. On the contrary, it is possible to extend, rather than to replace, the current theory of spatial database systems and GIS. Furthermore, moving from an exact to a vague domain does not necessarily invalidate conventional geometry; it is merely an extension. Consequently, the current exact object models that are restricted to determinate spatial objects can be

considered as simplified special cases of a richer class of models for general spatial objects. It turns out that this is exactly the case for the model to be presented.

The paper's structure is as follows, Section 2 presents modeling a fuzzy spatial datawarehouse considering the case when vague regions are included. Section 3 presents the construction of the fuzzy spatial data warehouse. Section 4 presents results and conclusion.

2. Related work

There are several representations and manipulation of vague regions, one of the most representatives is the Fuzzy Minimum Boundary Rectangle (FMBR) [4], that includes use of Fuzzy Logic to define degrees of memberships according to a membership function. In many geographical applications there is a need to model spatial phenomena not simply by sharp objects but rather through indeterminate or vague concepts. FMBR have been used to model such geographical data; it is considered an adequate tool to represent problems related with vague regions. FMBR is composed by two regions, the first region, called kernel, describes which part of the vague region belongs to it. The second region called boundary describes the fuzzy area of a vague region [4].

The Fuzzy Data Cube (FDC) [5] is a fuzzy multidimensional structure to query data in a darawarehouse using OLAP tools. It was initially defined for the sales problem, but in a different context, it is built by evaluating a membership function for each attributes stored in the data warehouse. The result is a degree of membership that is stored in the FDC. Part of this paper consists on to extend the definition of FDC considering a Spatial Database.

Working with the semantic of the attributes in a data cube is another approach that has not been widely considered. The representation of this model helps to draw conclusions with a higher degree of uncertainty [6], based on the classification performance in order to assist the decision support tasks, which has been an application of data cubes.

3. Modeling the fuzzy spatial data warehouse

Considering the previous work on section 2 and given that there is no a single existing model that integrates the concepts of Fuzzy Logic, spatial databases and data warehouses, our main goal is to develop a model considering such kind of integration.

The information under research requires a domain compressing the operational databases and other sources either internal or external [3], as it is shown in Figure 1. We assume that the data warehouse is represented as a multidimensional model [3], which represents the architecture of it. Figure 2 shows the proposed architecture of a Fuzzy Spatial Data Warehouse related with risk zones of Popocatepetl volcano.

In data warehousing context a *fact* represents the civil Protection Plan for natural disasters, and they are measured by demographic density and time to evacuate an area under threat. The information can be obtained by integration from each of the data warehouse's dimensions. Each dimension is organized by hierarchical levels; for instance de Space dimension is divided into state, region, county and town represented by their Fuzzy Minimum Boundary Rectangle (FMBR). As you can see this represents the levels of aggregation [3].

The treatment of spatial objects with indeterminate boundaries is especially problematic for the computer scientist who is confronted with the difficulties how to model such objects in a
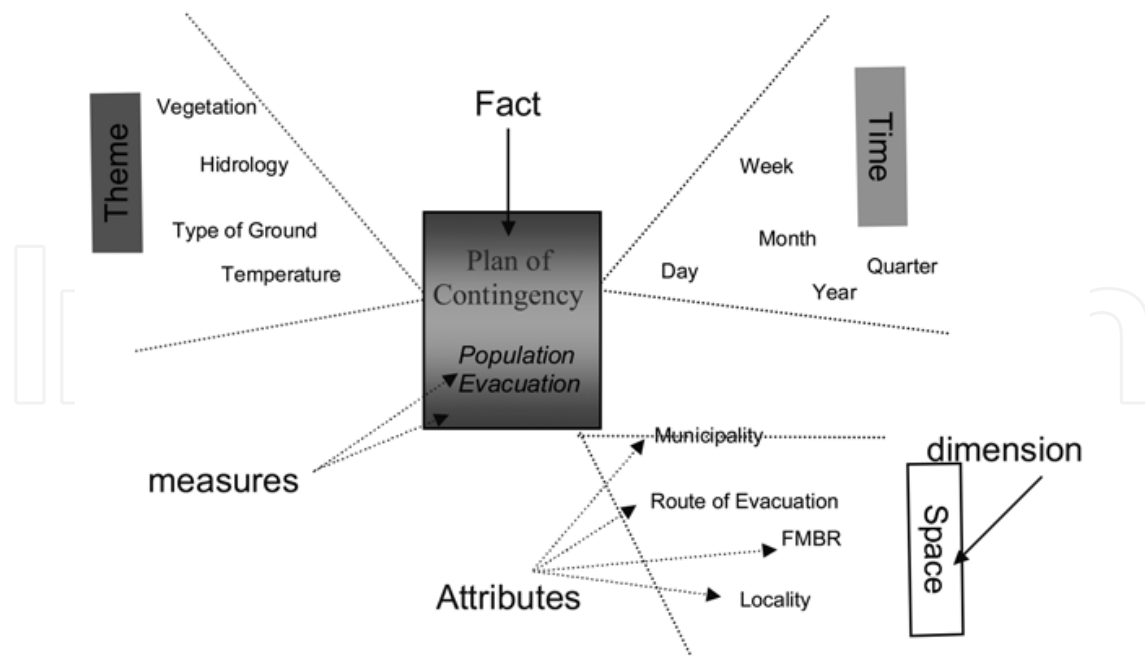
Fig. 2. Architecture proposed of a Fuzzy Spatial Data Warehouse

database system, so that they correspond to the user's intuition, how to finitely represent them in a computer format, how to develop spatial index structures for them, and how to draw them. Computer Scienctists are accustomed to the abstraction process of simplifying spatial phenomena of the real world through the concepts of conventional binary logic, reduction of dimension, and cartographic generalization to precisely defined, simply structured, and sharply bounded objects of Euclidean geometry like points, lines, and regions.

To define the Fuzzy Spatial Data Warehouse we begin with the dimensionality of the warehouse, based on the study case known, as risk areas nearby Popocatepetl volcano. To handle the information we use a Geographical Information System that contains georeferenced data from maps of the State of Puebla.

In Figure 4, it is shown the Snow Flake Schema [7] to be used in the Fuzzy Spatial Data Warehouse, as an extension of the multidimensional model shown in Figure 2. Notice how each of the dimensions are laid out. The main reason behind to select the Snow Flake Schema, is that the fact table is a table that uses the main table to relate with other fact sub tables. In such way, the datawarehouse has a tree like representation, where the root represents the principal fact table, and every node at the first level represents dimension and the remainder nodes with a level greater than 1 are called sub dimensional nodes.

Consider that a georeferenced data follows a recursive definition [11] of spatial concept, thus, it can be used to define a geographic concept. That is a geographic concept is either (i) a geographic data element, or it is (ii) a set of geographic concepts. Analogously, the physical manifestation of a geographic concept, namely a geographic object, is also expressible recursively as either a geographic element, or as a set of geographic objects, this definition is very important because that gives us the opportunity to work with star schema without ambiguity [7].

Thus far, working with the Snow Flake Schema it is possible to unfold each of the tables that each dimension describes, and adding the fuzzy spatial component to each table in order to be represented in the data warehouse.

3.1 Getting the spatial data from ArcGIS

The office in charge of natural disasters in México defined several risk zones in the Popocatepetl area, Figure 3 shows these areas including the county division in the State of Puebla. All the geographical information has to be stored in a vectorial format, where lines, points are stored as attributes of a spatial data base and each element represent a coordinate pair (x,y).
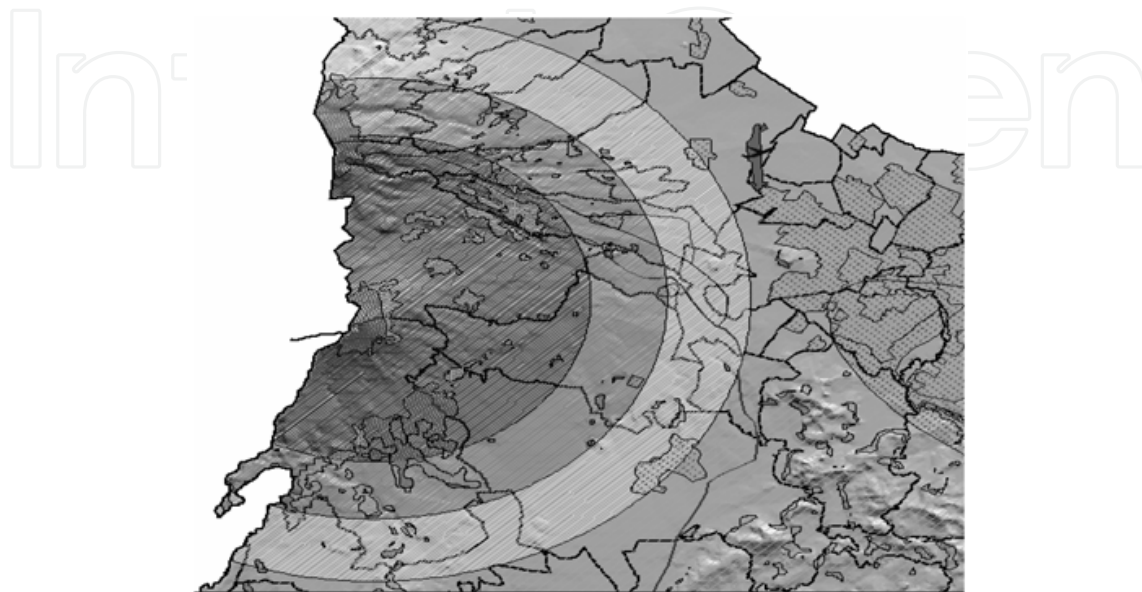

Fig. 3. Risk Zones in State of Puebla as they are shown with ArcGIS

Even though that the risk zones are well defined, they have a strong fuzzy component making that each object from a map (points, lines, polygons) to grow in volume, as it has been referenced, affecting the performance of the datawarehouse.

3.2 Fuzzy sets for the fuzzy spatial datawarehouse

Based on the nature of the geographic information, a fuzzy set is defined and related to each of the dimensions of the proposed schema (Figure 4). According to that schema for each dimension we have a set that represents certain degree of membership and based on that a label or a linguistic label is assigned. With that relationship, all the attributes get a semantic meaning in the warehouse (Figure 5).

Let see how that approach works; assume we have for each dimension or fact table a fuzzy set; and consider the fact table Plan of Contingency and we want to assign the degrees of membership to this set; notice that the available labels are MUCH, FEW or VERY FEW which represent the population where this variable belongs.

3.3 Spatial information in fuzzy spatial data warehouse

When researchers work with spatial data and data warehouses, they find that they know a few about implicit grouping hierarchy of such data, that is a problem because the preaggregation methods can not be applied to OLAP's operations [2][8][9]. Based on that, it is difficult to obtain different levels of grouping inside the data warehouse. In order to overcome this problem, three solutions are proposed in [2][8][9]:

1. Store the spatial pointers of each object without calculating the spatial measures from the spatial data cube.
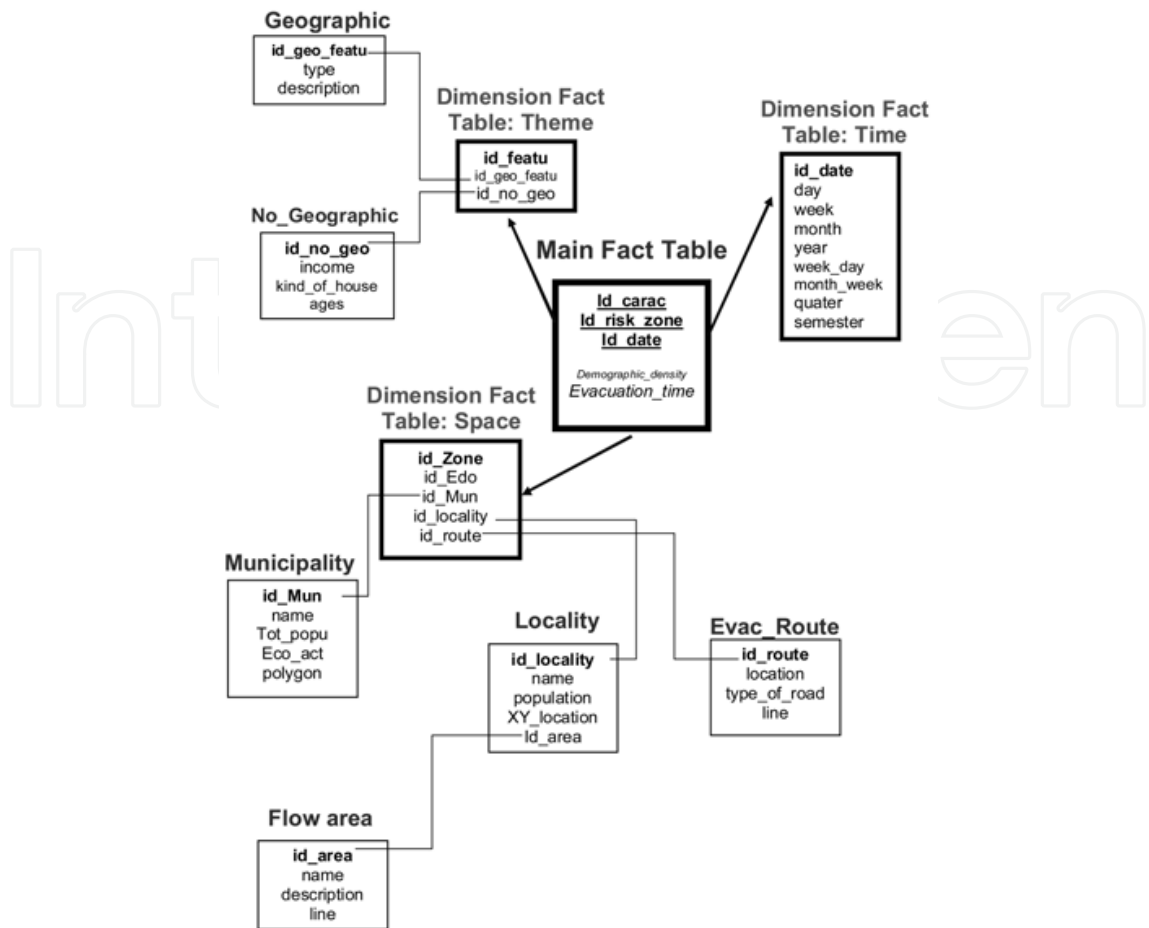
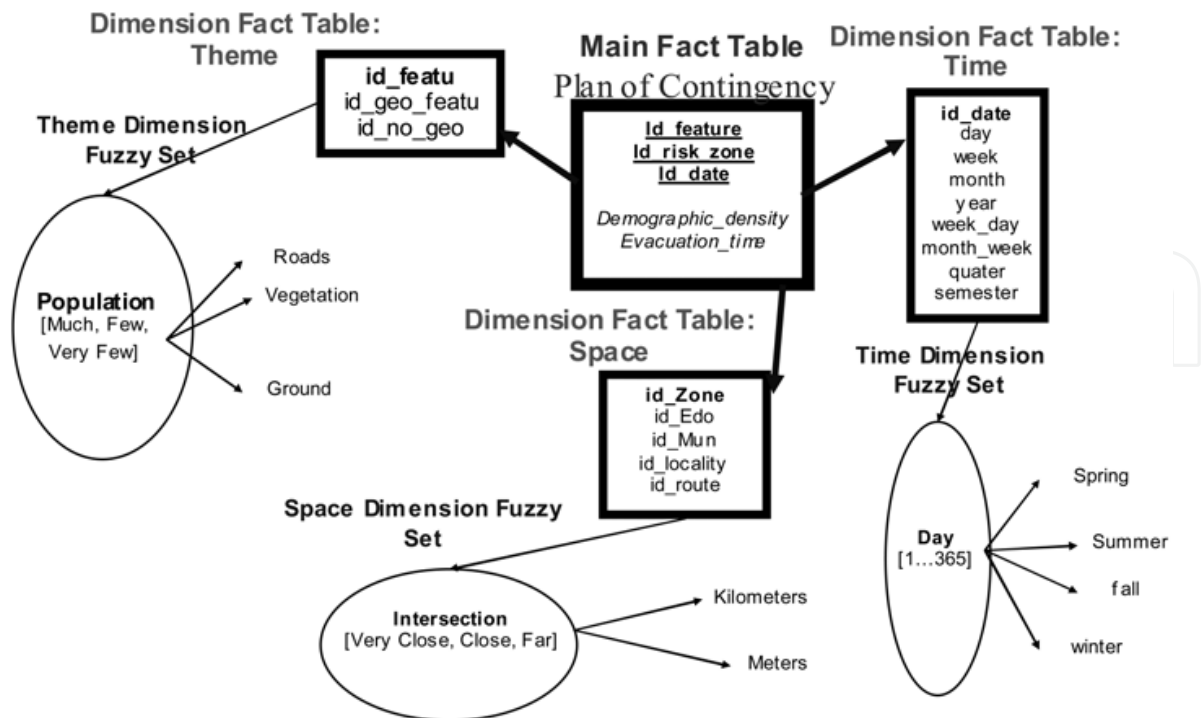Fig. 4. Snow Flake Schema

Fig. 5. Fuzzy Sets related to Data Warehouse Dimensions

2. Precompute and store some of the estimations about the spatial measures in the spatial data cube.
3. Perform selective operation measuring of some spatial measures the spatial data cube.

The way we obtain the spatial measures from the spatial data cube is by generating the information during the ETL process [10]. This process requires of spatial queries instead of query the data warehouse directly. We proposed to do this in three steps [10]:

1. Select the results that are close to the expected output from an SQL statement.
2. Load those results into the Data warehouse and create a new table for the spatial dimension.
3. Do the OLAP operations such as the generation of the data cube.

The first step is the most important, because it is possible to generate complex spatial queries [10]; together with the GIS, some of the operations that we perform are topological, intersection and distance operations among others. Figure 6 shows those operations involved.
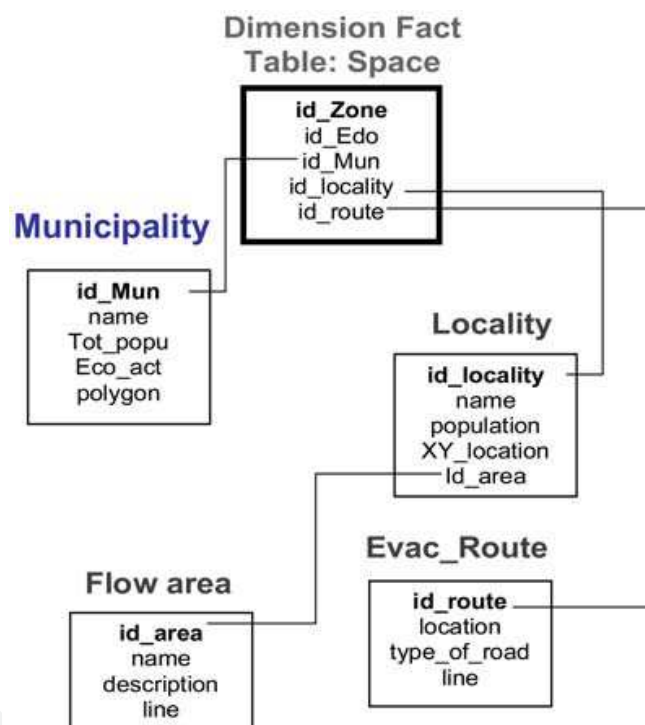

Fig. 6. Space Dimension from the Snow Flake Schema

For each table from Figure 6 is possible to create another spatial queries according to the information stored in the spatial database, and then, generate sort of aggregated data to be used in a future spatial query [2][10]. For instance, in County Table we have a polygon as spatial feature with its Cartesian coordinates.

As you can see in Figure 7, the attribute Polygon shows the coordinates for each county idenyified by ID_num, and working with this information inside the datawarehouse is not possible, because its representation is by levels of aggregation. If we want to answer a query like Q1: "Show all towns closer to an evacuation route", using the information stored in the ArcGIS's spatial database, what we can obtain is:

• The relationship between a distance value and the towns from the spatial fact table.
• Each distance value stored with a unique ID in order to distinguish them.

Id Mun	Name	Tot popu	Eco act	Polygon
1	ACAJETE	351	PLANT	(617350,2096890) ,(547660, 2135079),(651626, 2177004) ,(655233, 2177432),(617350,2096890) , (547660, 2135079)
2	ACATENO	423	FARMING	(675293, 2106590) ,(669084, 2706790),(679084, 2806293) ,(685233, 2101510), (675293, 2106590)
3	ACATLÁN	123	PLANT	(545869,2096890) ,(547660, 2135079),(595239, 2273004) ,(595999, 237732), (545869,2096890)

Fig. 7. County Table

- The attribute Polygon dropped in the County Table.
- A new table containing all the values.

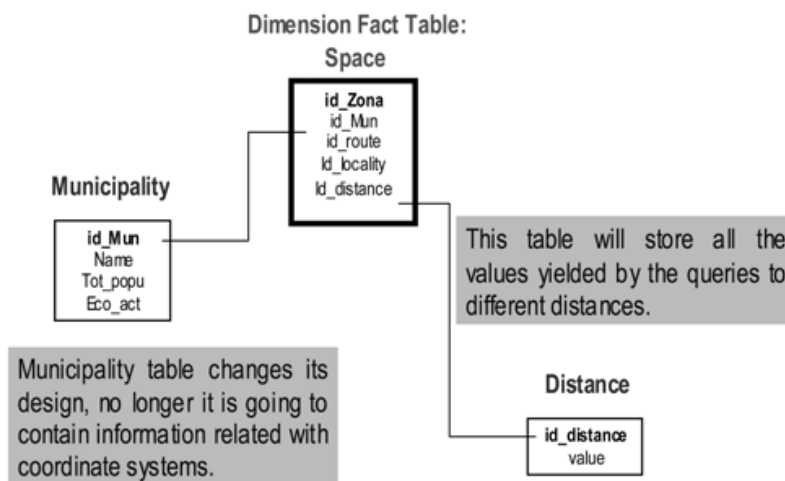Figure 8 shows such changes in the County Table.

Fig. 8. Changes to the Schema of County and Distances

Our model let us to have several types of queries based on differents points of view. In our study case we proposed four different queries (considering distance) related to the spatial dimension.

Figure 9 shows how Distance, Within, Intersection, FMBR Asociated will be the new tables to Spatial Dimension. Intersection deals with the level of danger related to a town, Within shows the number of inhabitants close to an evacuation route and FMBR Asociated considers a potential area of lava flow close to town(s). Based on the schema of the figure 9, it is how the Spatial Information can be represented inside the data warehouse.

3.4 Fuzzy spatial data warehouse membership functions representation

Once all the spatial values have been assigned to the warehouse, one task has to be done; that is, the calculation of the degree of membership to each register that will be stored in the table. For example, let's see the table Within, that table is related with the number of inhabitants that live close to an evacuation route and its linguistic values are Very Few, Few and Many. These values depend on the amount of people who live in the surrounding area of an evacuation route.
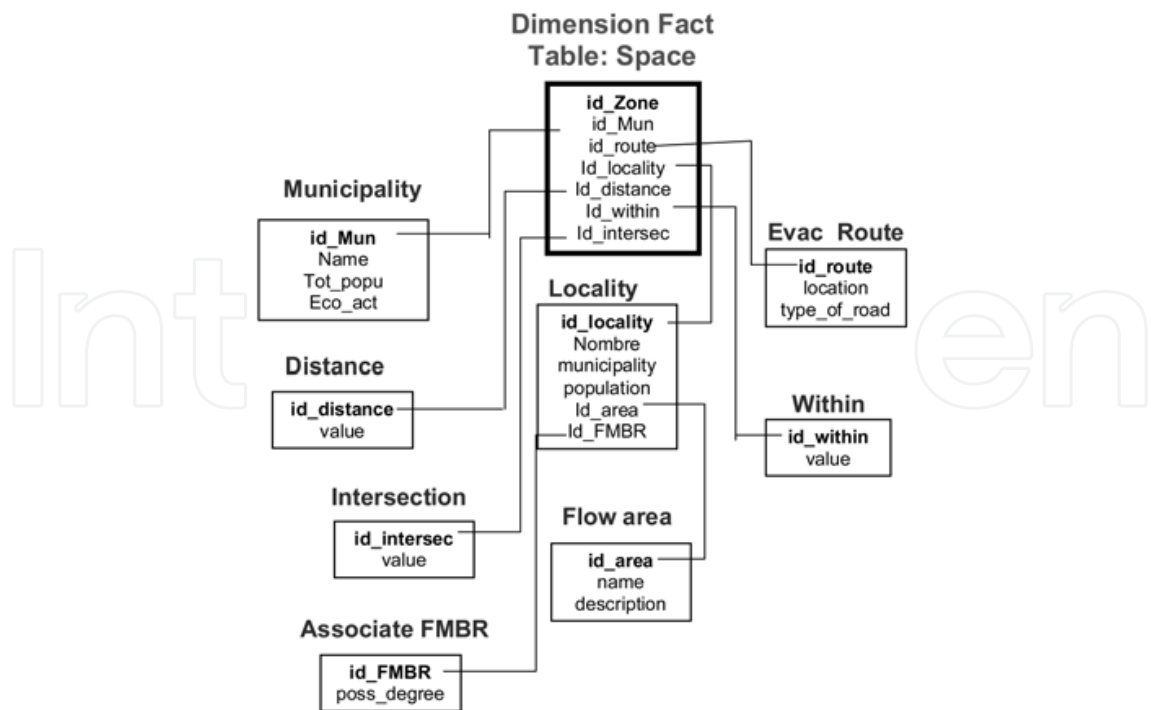
Fig. 9. Modified Schema for Spatial Dimension

As you can see in Figure 10 for an spatial value, a fuzzy value is matched, and its value comes as a result of evaluating the membership function.
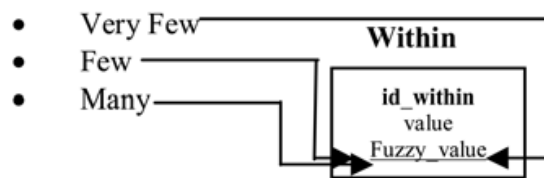

Fig. 10. Linguistic values associated to within function

In Figure 11 is shown the way an emprical value is calculated based on the Within Table, notice the relationship between the maximum speed limit in one road and the type of evacuation route and the number of passangers that certain bus can carry.

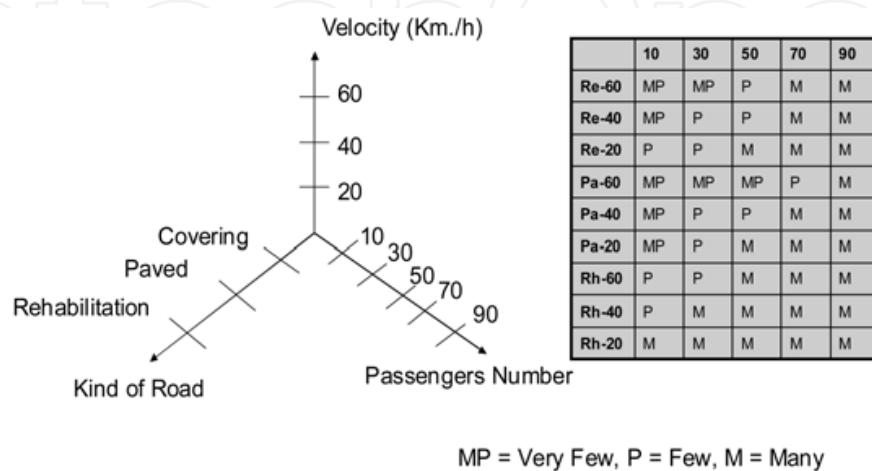|	10	30	50	70	90
Re-60	MP	MP	P	M	M
Re-40	MP	P	P	M	M
Re-20	P	P	M	M	M
Pa-60	MP	MP	MP	P	M
Pa-40	MP	P	P	M	M
Pa-20	MP	P	M	M	M
Rh-60	P	P	M	M	M
Rh-40	P	M	M	M	M
Rh-20	M	M	M	M	M

MP = Very Few, P = Few, M = Many

Fig. 11. Empirical Estimation of the Linguistic Variables

During the transition process the degrees of membership between speed, type of road and number of passangers are assigned according to the linguistic values, that variables have.

A new query needs to be done in order to determine if the statement could be fullfillment or not, such a query (statement) is *"E1= Traveling in a paved road at high speed and carrying very few passangers"*. Table 1 shows the results of the evaluation of the command E1through the evaluation of the aggregated membership function (1). The result tells us that does not matter how well are the conditions, if the number of passangers is very low and the required time is not an issue.

$$\min(\max[\mu(road), \mu(velocity), \mu(passengers)]) \tag{1}$$

	Road	Velocity	Passengers Num.
Pa-high-MP	1	1	0.2
Pa-half-MP	1	0.6	0.2
Pa-low-MP	1	0.3	0.2
Pa-high-P	1	1	0.5
Pa-half-P	1	0.6	0.5
Pa-low-P	1	0.3	0.5
Pa-high-M	1	1	1
Pa-half-M	1	0.6	1
Pa-low-M	1	0.3	1

$$\min(\max[\mu(road), \mu(velocity), \mu(passengers)]) = 0.3$$

Table 1. Evaluation of the command E1

Kind of Road	Paved	Covering	Rehabilitation
μ(Road)	1	0.6	0.3
Passengers Number	From 56 to 90	From 36 to 55	From 10 to 35
μ(passengers)	1	0.5	0.2
Velocity (Km/h)	60	40	20
μ(velocity)	1	0.6	0.3

Table 2. Evaluation of the linguistic variables on (1)

Table 2 shows the empirical assignments of the membership degrees of the linguistic variables that are involved in the command E1. Based on the values assigned to the linguistic variables, the degree of fullfillment of E1 was found 0.3. This result means that is very few probable than few passengers be traveling in a paved road at high speed.

4. Conclusions

This work represents a part of the Intelligent Geographical Project, that will model a geographic area in the same way as in the real world appears, taking advantage of the

Information Technology. We have integrated ArcGIS technology with Fuzzy Theory, given the fact that linguistic variables get closer to coloquial language and they describe a geographic situation in a natural way.

The main contribution of this work is the integration of Fuzzy Logic with Spatial Databases in order to help during the decision support and OLAPs querying processes.

ArcGIS allows to obtain spatial features by the queries execution on maps. These spatial features are integrated into a multidimensional database allowing aggregation and disaggregation OLAP operations on it, avoiding the use of classic spatial access methods. In addition, spatial semantic is added to spatial and not spatial dimensions of the multidimensional model improving the decision making process. Finally, the Fuzzy Spatial Data Warehouse's design methodology proposed, simplify the use of the existing analysis tools for exploting the potential of Data Warehouses.

5. References

[1] Y. Wang, H. Shao. Data warehouse technology in process industry. in Intelligent Control and Automation. Proceedings of the 3rd World Congress. 2000.

[2] D. Papadias et. al. Efficient OLAP Operations in Spatial Data Warehouses. in Lecture Notes in Computer Science, vol. 2121. p. 443 - 459. 2001.

[3] J. Hernández, J. R. Quintana, C. Ferri. Introducción a la Minería de Datos. Capítulos 1, 2 y 9, Pearson Prentice Hall. 2005.

[4] M.J. Somodevilla, Fuzzy MBRs Modeling for Reasoning about Vague Regions, Doctoral Tesis, Tulane University, 2003.

[5] A. Reda, K. Mehmet, "Integrating Fuzziness into OLAP for Multidimensional Fuzzy Association Rules Mining", Third IEEE International Conference on Data Mining (ICDM'03) 2003 p. 469.

[6] L. Yubao, Y. Jian, "The Computation of Semantic Data Cube", GCC 2005, pp. 573-578.

[7] M. Levene, G. Loizou. Why is the Snowflake Schema a Good Data Warehouse Design? in Source, Information Systems Vol. 28 (Issue 3). pp. 225-240. ISSN 0306-4379. 2003.

[8] H. Ahn, N. Mamoulis, H. Wong. A Survey on Multidimensional Access Methods. UU-CS. 2001.

[9] N. Stefanovic, J. Han y K. Koperski. Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes. in IEEE Transactions on Knowledge and Data Engineering, Vol. 12, No. 6. 2000 p. 938-958

[10] Y. Li, Y. Chen y F. Rao. The Approach for Data Warehouse to Answering Spatial OLAP Queries. In Intelligent Data Engineering and Automated Learning. p. 270-277. SpringerLink Date Tuesday, August 26, 2003. ISBN 978-3-540-40550-4.

[11] Aldridge, C.H., 1998, A Theory Of Empirical Spatial Knowledge Supporting Rough Set Based Knowledge Discovery in Geographic Databases. Ph.D. Thesis, University of Otago, Dunedin, New Zealand.

Decision Support Systems Advances in

Edited by Ger Devlin

This book by In-Tech publishing helps the reader understand the power of informed decision making by covering a broad range of DSS (Decision Support Systems) applications in the fields of medical, environmental, transport and business. The expertise of the chapter writers spans an equally extensive spectrum of researchers from around the globe including universities in Canada, Mexico, Brazil and the United States, to institutes and universities in Italy, Germany, Poland, France, United Kingdom, Romania, Turkey and Ireland to as far east as Malaysia and Singapore and as far north as Finland. Decision Support Systems are not a new technology but they have evolved and developed with the ever demanding necessity to analyse a large number of options for decision makers (DM) for specific situations, where there is an increasing level of uncertainty about the problem at hand and where there is a high impact relative to the correct decisions to be made. DSS's offer decision makers a more stable solution to solving the semi-structured and unstructured problem. This is exactly what the reader will see in this book.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Perez David, Somodevilla Maria J. and Pineda Ivo H. (2010). Fuzzy Spatial Data Warehouse: A Multidimensional Model, Decision Support Systems Advances in, Ger Devlin (Ed.), ISBN: 978-953-307-069-8, InTech, Available from: http://www.intechopen.com/books/decision-support-systems-advances-in/fuzzy-spatial-data-warehouse-a-multidimensional-model

INTECH
open science | open minds