EXPERT SYSTEM TO CLASSIFY SEMANTIC INFORMATION TO IMPROVE MAP DESIGN

Msc. Marcio Augusto Reolon Schmidt[1]
Prof. Dra. Luciene Stamato Delazari[2]

[12]Programa de Pós-graduação em Ciências Geodésicas
Departamento de Geomática
Universidade Federal do Paraná
Curitiba BRAZIL 81531-990
marcio.schmidt@gmail.com, luciene@ufpr.br

Abstract – This article presents the implementation of an expert system based on rules to classify semantic information in order to assist *Secretaria de Estado do Trabalho, Emprego e Promoção Social (SETP)* at Paraná, Brazil. This bureau is responsible for definition of social assistance policies and actions for public welfare. Despite the problem of poor quality maps built by casual users due to the popularization of GIS software, situation that has been discussed in cartography for about 20 years, this problem still occurs in Brazil. That was the motivation for this article. The problem seems to be related to classification of information concerning levels of measurement and, therefore, on how to symbolize maps. The situation occurs often at municipal departments, NGOs and other institutions, using maps for planning, support decision making, and happens, at least in part, because of ease to use GIS software and lack of training for institutions employees. In this context, an expert system seems to be a better option to ensure casual users can make right choices in map making process. The expert system presented in this article is embedded on a CIS (Cartographic Information System) acting as a data manager, using IF-THEN rules. These rules identify which kind of information is being demanded to be presented, and makes its classification. In its current stage, CIS allows semi-automatic classification of thematic information in nominal and ordinal levels of measurement, and allows users to manipulate data and its representations under criteria of map design. This representation provides transmission of information at the same time, so it's safe to say information will be properly processed in database, without user intervention. As a major advantage, system exemption the user training in subjects not essential to exploration of spatial data.

1 – Introduction

In this article paper we present the experience of implementation of an expert system to classify thematic semantic information. The expert system is assisting *SETP*, at Paraná, Brazil.

The article was motivated by problems seen in maps produced by casual users of GIS. Despite the software popularization, it's not safe to say non-cartographers users can make maps under its design rules. Although is desirable users produce their own maps, these maps do not fulfill users needs when their goal is to make space analysis, due to poor quality of maps created. This happens because GIS software became widely available and accessible to a community of casual users without knowledge in cartographic representation principles. Since the knowledge of map design is no longer exclusive of cartographers, according to (Robbi, 2000; Weibel & Buttenfield, 1992), ill-structured maps, with several errors of design, are seen in newspapers, scientific literature, on the internet and decision support systems.

It's usual to notice it how municipalities and NGO use and generate maps for planning cities and carry out social and ecological analysis. The problem in Brazil, by our point of view, is concentrated on how to classify information in terms of level of measurement and, therefore, how to symbolize maps, the final steps of map design.

Several authors pointed out the need for researches about how to assist users on map designing maps with GIS. For automatic visualization, Casner (1991), Roth et al (1994) and Senay and Ignatius (1994) developed studies for scientific visualization and geovisualization under generation and testing paradigms. Those studies originated systems aiming for the elimination of the need to specify, design and develop different visualizations for each outcome of a program, like GIS, allowing users to focus on their attention more appropriately (determining and describing information to be represented). Recently CommonGIS (Andrienko & Andrienko, 2003) and Geoda™ (Anselin, 2004), were presented, but their focus is on HCI, an interactive training assistance and EDA assistance tool. CommonGIS focus on adapting interfaces as users explore the tool and acquire system knowledge. The exploration is guided by an expert system giving users hints and options. Geoda™ has its emphasis on datamining and spatial statistics, and it includes functionality ranging from simple mapping to exploratory data analysis, using interactive environment combining maps with statistical graphics (Anselin, et al, 2004). However, there are still problems about data classification and symbolization.

2 – Personal Map and Casual User

Robbi (2000) points that maps construction made by casual users have basic errors of design. It’s a consequence of technology fast evolution, resulting in computer users with minimal or even nonexistent knowledge about cartography. Referring to Kolacny’s model, insertion of users on map making process shifts cartographers out of process, restricting its role to foresee how GIS would be used. There is a break at the setting up context of map design because cartographers and users do not share the same “reality”. Map designers and users become the same and, as a result, it’s possible to say maps are made for personal use (Van Elzakker, 1999), since their classification, representation methods and colors are chosen toward aesthetic purposes; communication remains almost incidental. Van Elzakker & Wealands (2007) point that variations of using multimedia map making are great, as well as the variety of users ranging from basic to professional. This is especially important for the process of design in these systems. After conceive which information will be used to fulfill users needs, map makers should go for the process of data classification. This is a fundamental aspect, because users experience on reading maps under different classification methods may vary intensely. This is a detail with direct influence on symbolization.

Symbolization, as defined by Bertin and spread by others eminent cartographers, is a combination of visual variables, graphical primitives, and classification of data in scale of measurement. This article refers directly to the scale of measurement on choropletic maps (on categorical data). According to Chrisman (2008, p.381; 1996 p.9), the scale was proposed by Stanley Stevens in 1946 as a scheme to organize any measurement into a grouping, to attribute them common properties. Stevens' scheme presented four classes of measurements, called "levels" or "scale" of measurement: nominal, ordinal, grouped as categorical scale of measurement, and numerical, divided in interval and ratio scales.

Understanding these concepts into the process of map design is a major problem for casual users and can be noticed in Brazil nowadays. It's not unusual to find maps with continuous color ramps representing discrete data, map projections problems, complex symbols with no important information, and "noisy" visualization making any trying of interpretation almost impossible. This situation occurs often at municipalities departments, NGOs and institutions using maps to planning support decision making process, because of popularization of GIS software use and lack of training of employees or institutions staff.

Questions are: how to help casual users to classify attributes and symbolize maps according to the principles of map design? If casual users cannot be trained, how is it possible to ensure they will achieve minimal level of understatement about correct use of GIS? If there are no map design principles for digital environment, mainly because cartographers do not know exactly which points can be adopted from traditional map design theory, which considerations should be taken into account to make these principles feasible?

3 – Cartographic Information System for Spatial Data Classification

We deal partially with the question presented previously. Symbolization occurs in terms of level of measurement, graphic primitives and visual variables. Casual users do not always understand the meaning of these concepts and does not know how to select elements of map symbols at final steps of map designing. There are several possibilities to ensure casual users will make right choices, ranging from expensive time and money long-term training to other options, like artificial intelligence systems. Expert systems seem to be a better choice.

Andrienko et al (2003) justified use of an intelligent guide to help users to employ properly the instruments. According to him, new techniques should be introduced to them when their own analysis is performed. Authors make a list of requirements that guiding systems had to satisfy: recognition of users current goals (tasks to be accomplished), and giving users essential explanations to understand how to use instruments properly for each task. Authors performed tests with users to evaluate learning process of tool use, and suggested that short training, associated with previous demonstration, enable people to use tools locally installed. When no direct teaching is possible, substitutes need to be provided. The authors also point that online help or interactive online tutorials are not always effective, because people will not invest time and effort in such training without being sufficiently convinced of new tools benefits.

However, Andrienko et al (2003) wrote about an expert system directed to users trained on interactive tools usage. It is hard for a casual user to comprehend concept of level of measurement and its consequences on map design. Expert systems should run under the interface taking care of user actions, managing information and considering map design aspects, to avoid mistakes. As a result, it's possible to quote Van Elzakker (1999), who explained the need of cartographic expert systems to enable handling map design and its user aspects. For this article it was incorporated an expert system to attribute classification into a stand alone Electronic Atlas, called *Social Atlas.*

3.1. - Understanding Users

Originally, the objective was support the *SETP*, responsible for define social assistance public policies and its implementation. In Paraná, this bureau acts as a governmental manager, defining destination for financial resources. *SETP* technicians, in 2000, needed to know how counties were organized in terms of Municipal Councils and Public Funds, in order to implant the social law lately approved at that time. For that reason Delazari (2004) started to work on an electronic Atlas prototype, called *Social Atlas*. The objective of *Social Atlas* was to carry out space analysis, consulting database and generate maps by interaction, based on the needs of this bureau under the context of LOAS (Lei Orgânica de Assistência Social).

The *Social Atlas* has passed for many revisions and improvements since the first edition. Nowadays, the system can generate 26 choropletic maps, and integrate an expert system that manage and classify semantic information in terms of level of measurement, subject of this paper. For these reasons it is justified consider that *Social Atlas* is, in fact, a Cartographic Information System (CIS), as described in Pucher et al (2007), Shulei, Yufen (2007) and Yufen,Yun (2007). To the authors, the recent existing atlas information systems and others kind of CIS based on multimedia electronic atlas (case of *Social Atlas*), desktop GIS, web maps and mobile maps show its development towards a cartographic information system (CIS). The main characteristics of CIS are no clear distinction between non-specialist and expert user. Such systems take care of individual user abilities and skills by adapting themselves to their needs.

3.2. – Implementing the expert system for data classification

An expert system placed between the database and the representation device was built. This system manages database through application of production rules described as IF-THEN. This expert system connects database developed in Access and the representation device, implemented with MapObjects (ESRI), controlling flow of data in two situations: insertion of new data and the consultation through SQL. Production rules are suitable to problems in which knowledge is well established and can be easily described as IF-THEN rules (Rodrigues-Bachiller, 2004, p.42). Those systems needed a set of facts to represent initial work memory, a set of rules that handle any aspects of the problem, and a condition to inform to the system if the problem was solved or not.

Initial memory, or initial facts, is all information located in database, collected by technicians of SETP bureau to implementation of LOAS. Data is organized in three themes, or major groups, which separate information in terms of its characteristics. Each theme has several classes and its condition is related to define data´s level of measurement and knowledge base rank. The set of rules, differently from other expert systems, is embedded in the software, and the parameters of those rules are updated by use, and can be accessed by the user. The set of rules try to indentify which kind of data it's been inserted. Expert system tests type of data to insert new information. Numerical data is classified by numerical level. *Social Atlas* does not distinguish between intermission or ratio level of measurement, because map design does not consider it. Nominal or ordinal data are defined based on ordering of knowledge base. Nominal and ordinal data are stored at the same knowledge base, but can't be mixed because of data nature. Therefore the rules and representation are quite different. The next step is indentifying whether the information is already known. The set of rules searches at knowledge base by an order associated to category names.

When dealing with non numeric data, e.g., semantic classification, the choice for one or another level of measurement demands attention, because correct order is subjective. Data indicating temporality or any kind of order can be considered as ordinal data and need to be evaluated in details. For example, for LOAS implementation it is important to know County Counsels creation data. Classes are "first semester of 1995", "before 1995", "second semester of 1995", "after 1995" and "no available data". Expert system searches the whole class, e.g. first semester of 1995, for possible ordering of categories. When ordering is found to all categories, rules select the visual schemes (color ramps) adequate to representation. Figure 1, shows an extract of the algorithm.

```
Data kind:
If Number Then
        Level of Measurement = Numeric → Enable visual variables

If NOT Number Then
        If Consult Knowledge Base has Order indication Then
                Level of Measurement = Ordinal → Insert ordering
        If Consult Knowledge Base has NOT Order indication Then
                Word Wrap class names
                If Dictionary has Order indication Then
                        Level of Measurement =Ordinal → Insert ordering and Enable visual variables
                If Dictionary has NOT Order indication Then
                        Ask user to confirm there is NOT Ordering
                                If Confirmed
                                        Level of Measurement = Nominal → Enable visual variables
                                If NOT Confirmed
                                        Try again or Ask user to indicate order
END
```

Figure 1. Extracting algorithm

As seen in figure 1, above, if it's not possible to be defined, categories names are broken into a list of words with word-wrap function. Position of each word in sentence is stored. This list is compared to knowledge base. Inside there's an associated dictionary. This dictionary keeps words in a more generic sense, giving local and global ordering of simplest part of class names. The dictionary is composed of words and prepositions, and used to classify words presented in the list. Words like "Until", "Before", "Between", "Among" and "After", are part of the dictionary, also working as a memory, keeping all words used for classification. The dictionary is also an excellent resource for present and previous classification, so if a class has had been deleted and a new one with a similar name is inserted, the system is able to estimate a possible place for the new class. As seen previously, if all class names cannot be found, system will try to classify the first word of each class, and then the second word of each class, and so on, trying to establish a hierarchical relationship between class names. If it's successful, knowledge base feedbacks itself with new classification. In case of failure or partial success, users could be inquired to indicate correct class order. Then, the system stores this user classification, feeds the dictionary and carry on with representation. As users supply knowledge base with more information, vocabulary becomes more extensive and system can deal with complex situations, so human intervention becomes less necessary. This is interesting when an experienced technician built knowledge base and dictionary, and distributed them with the *Social Atlas.* Figure 2 shows a sketch of how system works.

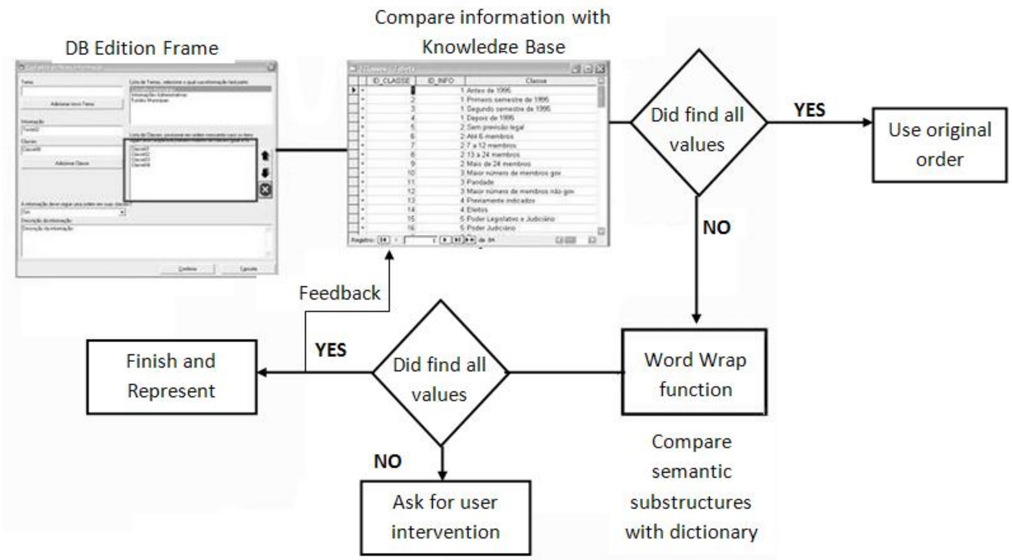

Figure 2. Expert System Sketch

The figure 3, above, presents the interface of expert system. This dialog box is accessed from menu Edition of *Social Atlas*. All other steps of expert system run under this dialog box and users do not get in direct contact with data or its classification.

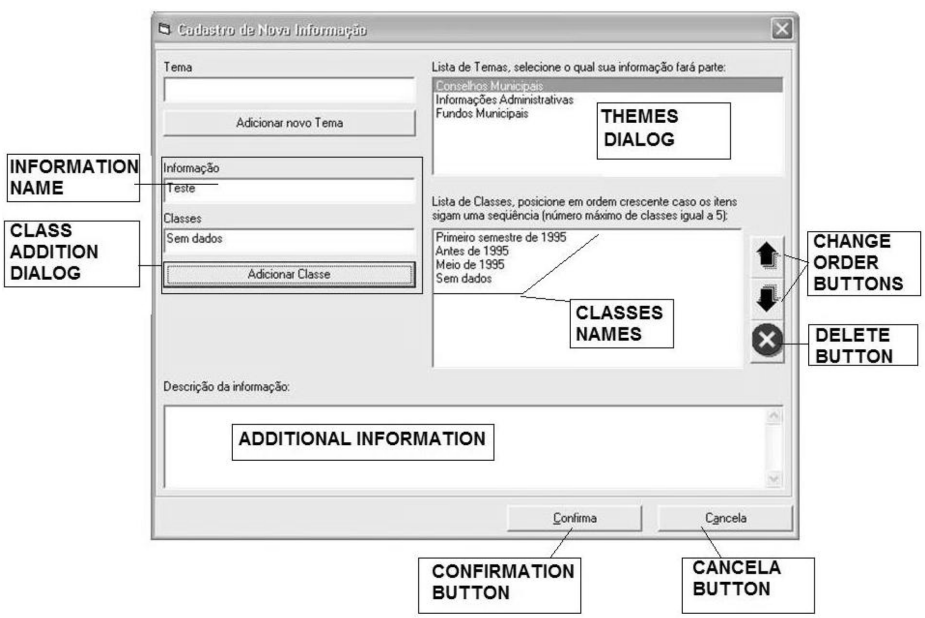

Figure 3. Expert System Interface

The themes are shown in Themes Dialog Box. Data, i.e. column name, is supplied by the class dialog box on the left. In this dialog box the class names are supplied and appear at the right side. Clicking on ‘confirm’ button makes the system carry on with classification, as described previously. In case of no success, right side buttons are enabled and a message box pops up asking the user for intervention, as explained before. Users sometimes require storage additional information as text. This action can be done in the field Additional Information field, at the bottom of the dialog box.

4 – Results and evaluation process

The system evaluation was performed considering two different groups of users who answered the same questionnaire. First group consisted of 5 cartographers, or people with prior knowledge of mapping software. The second group consisted of 6 professionals of Social Assistance, with very little or none experience of map making. A set of tasks using Social Atlas and the expert system was defined. At the end of each task users should answer "yes" or "no" to some questions. At the end of the test users answered a questionnaire for global assessment of the system.

A qualitative analysis of results was performed considering expertise of groups evaluated. Social Service professionals were analyzed for their comprehension, use of functions and difficulty found during implementation and assessment of data classification. Cartographers answered the same questionnaire, showing good comprehension of functions and implementation structure.

Some responses given by non-cartographers users revealed system is not operated properly because of lack of experience with map making software. Answers revealed the prototype is easy to operate. Other important aspect is that users with previous map making experience expect to be able to manipulate representation of space information, such as change colors for classes in theme maps. non-experienced users are more concerned of thematic data and about functions that allow access to it, for example the functionality of editing texts of thematic maps

The actual knowledge base has correctly classified about 90% of data inserted. Errors remain in classes which, in Portuguese, have ambiguous sense and more than only one classification. This problem will be solved as users expand the knowledge base through use. Also, CIS can generate maps on assisted way considering nominal and ordinal levels of measurement.

System has been well evaluated by users and show it’s possible to use artificial intelligence in this context. Other important result is the consistency of CIS database, because data is now correctly classified. Map symbols are then controlled, avoiding possibility of selection of a continuous color ramp to represent nominal data, for example.

5 – Outlooks and Future Works

The system is being continuously improved. CIS is being converted to Geoserver, a Java-based map server. In this new implementation, the experiment was connecting the original knowledge base, stored in UFPR server, and the user-defined knowledge base, stored at their personal computers. Expert system is required to manage appropriately the data flow, identifying where data request will be represented. To connect information of both databases, a message should be sent to map server through a written web script. The script uploads classification and symbolism defined in CIS, sends a new message to users computers to upload user database and do the classification. After that, script writes the SLD file, e.g., a XML file that defines how data will be displayed at users computer screen. The new implementations are expected to be functioning by the first semester of 2009.

6 – Acknowledgment

To Conselho Nacional de Pesquisa (CNPq) by the scholarship and support.

7 – References

Andrienko,G., Andrienko,N. & Voss, H. 2003 GIS for Everyone: The CommonGIS Project and Beyond. In: Peterson, M. P. *Maps and the Internet*. Elsevier Science B.V.

Anselin, L. Syabri, S. & Kho, Y. 2004. GeoDa™: An Introduction to Spatial Data Analysis. Spatial Analysis Laboratory. Department of Agricultural and Consumer Economics. University of Illinois, Urbana-Champaign. Available at: https://www.geoda.uiuc.edu/documentation/publications [Accessed 15 November 2008]

Casner, S. M. A. 1991. Task-Analytic Approach to the Automated Design of Graphic Presentations. In: *ACM Transactions on Graphics (TOG) archive*. Volume 10 , Issue 2, p. 111 – 151. ACM Press New York, NY, USA. ISSN:0730-0301 Available at: http://doi.acm.org/10.1145/108360.108361. [Accessed 21 may 2007].

Chrisman, N. 2008. Scales of Measurement. In: Kemp, K,K. 1° Ed. *Encyclopedia of geographic information science*. ISBN 978-1-4129-1313-3 (cloth). California: SAGE Publications, Inc.

Chrisman, N. 1996. Exploring Geographic Information Systems. ISBN 0-471-10842-1. United States: JohnWiley & Sons, Inc.

Delazari, L. S. 2004. Modelagem e implementação de um Atlas Eletrônico Interativo utilizando métodos de visualização cartográfica. 155p. Thesis. São Paulo, Brazil, Escola Politécnica da Universidade de São Paulo, Departamento de Engenharia de Transportes.

Pucher,A.,Kriz, K,. Katzlberger,G.2007. Örok-Atlas Online – Atlas Information System Of Austria. Proceddings: 23th International Cartographic Conference (ICC2007), Moscow, Russia.

Robbi, C. 2000. Sistema para visualização de informações cartográficas para planejamento urbano. Thesis. 369f. São José dos Campos, Brazil.Instituto Nacional de Pesquisas Espaciais, INPE.

Rodrigues-Bachiller, A. 2004. Expert Systems and Geographical Information Systems for Impact Assessment. pp.400. London: Taylor & Francis Inc. ISBN 0- 415- 30724-4

Roth, S. F, Kolojejchick, J., Mattis, J., Chuah, M. C., Goldstein, J., Juarez, O. 1994. Sage Tools: A Knowledge-Based Environment for Map designing and Perusing Data Visualizations. Conference on Human Factors in Computing Systems. Boston, Massachusetts, USA. P. 27 - 28 ISBN:0-89791-651-4. Available at: http://doi.acm.org/10.1145/259963.259992, [Accessed 21 may 2007].

Senay, H. & Ignatius, E. 1994. A Knowledge-Based System For Visualization Design Computer Graphics And Applications. IEEE, Volume 14, Issue 6, p. 36 – 47, doi:10.1109/38.329093 Available at: http://ieeexplore.ieee.org/iel1/38/7784/00329093.pdf, [Accessed 21 may 2007].

Shulei,Z., Yufen,C. 2004. The Principles of Map designing CIS-Cartographic Information System. Proceeding of the 12th International Conference on Geoinformatics. P.389-P.396. Available at: http://www.fromto.itb.hig.se/~bjg/geoinformatics/files/p389.pdf [Accessed 30 June 2009]

Van Elzakker, P.J.M. C. 1999. Thinking aloud about exploratory cartography. In: C.P. Keller, *Touch the Past, Visualize the Future.* Proceedings 19th International Cartographic Conference, Ottawa, Canada. 1, pp. 559-569. Ottawa: Organizing Committee for Ottawa ICA 1999. ISBN 0-919088-54-6.

Van Elzakker, C.P.J.M. & Wealands, K. 2007. Use and Users of Multimedia Cartography. In: Cartwright,W. Peterson, M.P., & Gartner, G. 2º Ed. *Multimedia Cartography*. ISBN-10 3-540- 36650-4. New York : Springer.

Weibel, R.; Buttenfield, B.P. 1992. Map design for geographic information systems. *International Journal of Geographic Information Systems*, 6, p. 233-245.

Yufen,C.,Yun,L. 2007. Design And Implementation Of Multi- Style Gui For Cartographic Information Systems. Proceddings: 23th International Cartographic Conference (ICC2007), Moscow, Russia.