VisMade: Visualizing Temporal Information in Databases Structured as Master-Details

Edécio Fernando Iepsen[1,2], Paulo Roberto Gomes Luzzardi[1], Stanley Loh[1]

[1]Programa de Pós Graduação em Informática – Universidade Católica de Pelotas
Rua Félix da Cunha, 412 – 96010-000 – Pelotas – RS – Brasil

[2]Curso Superior Tecnologia em Sistemas de Informação – Faculdade Tecnologia Senac
Rua Dom Pedro II, 901 – 96010-300 – Pelotas – RS – Brasil

edecio@terra.com.br, {luzzardi, loh}@atlas.ucpel.tche.br

Abstract

Information Visualization is a field of computer graphics which aims to present information graphically, allowing users to better understand a great amount of information. It also allows identifying standards, or even detecting new information which, at first, was not possible to be perceived due to its complex representation. This study briefly describes this field and presents a tool for visual data mining executed on the web, through of the Parallel Coordinates technique, that allow users to analyze information structured in a master/detail format. Other advantage of the tool is its capacity of accepting data from several systems exported by the user in the XML format. A case study is demonstrated in order to exemplify the resources of the tool.

Keywords: Information Visualization, Master-Details Databases; Visual Data Mining

1 Introduction

The amount of information stored in electronic media has increased significantly over the last years. Several studies point out this fast growing [11] [18] [21]. However, having a great volume of information not always means the real possibility for enlarging someone's knowledge. For example, imagine the thousands of records in a database about admissions and treatments of patients in a medical care center; basic information as a report about patients admitted by month can be quickly obtained, however analyzing details as how many patients admitted in the begin of the month with a certain disease is a difficult task. We need mechanisms able to represent this huge amount of information in a more accessible way to human perception. This is the scope in which the area of information visualization works. By means of representations or interactive visual metaphors it aims to ease data comprehension of a given system.

This paper aims to present the **VisMade** tool, used for temporal information visualization stored in databases organized in the "master/detail" format, or through "one-to-many" relations. A lot of data are structured in this format; for example customers-sales, patients-admissions, hospitals-patients, physicians-patients, etc.

The organization of this work comprises a brief description of the areas involved in it (chapter 2). Next, in chapter 3, it describes the **VisMade** tool, developed through the technique of Parallel Coordinates. Chapter 4 shows a study case using the tool and the paper ends (chapter 5) with concluding remarks and indications of future works.

2 Information Visualization

Information visualization is an area of computer graphics which aims to generate images that ease human understanding especially on great volumes of data. Combining aspects of computer graphics, man-machine interface and data mining [2], [1], [7], information visualization allows data presentation in graphic mode, in order to facilitate the analysis and understanding of complex information. [3], [9], [18]. In this respect, it is important to emphasize that human capacity to deal with visual information is much higher than with textual data [21].

2.1 Information Visualization Techniques

Information Visualization techniques aim to graphically represent data through a way explores human capacities allowing him/her to interpret and understand information and to deduce new knowledge [5].

The information visualization techniques already developed use representations or visual metaphors to graphically exhibit data which do not usually have

direct, obvious, and natural representation. In several techniques, authors often search for inspiration in real world objects (or geometric ones) for mapping the set of information. The visualization techniques may use visual representations 1D, 2D, or 3D, not necessarily in accordance with information space dimension [13].

The choice of an Information Visualization technique for a specific set of data depends on several factors. Just as in the process of a graphic creation in worksheet programs, in which several formats for the graphic creation are presented, the building of tools for information visualization follows the same logic. Examples of Information Visualization techniques are (illustrated in figure 1):

a) **Focus + Context**: it allows exhibiting a set of information to the user, where he can point out the part of his interest – usually amplifying the selected piece of information [17]. The techniques of the focus + context type are also known as *Fisheye*, exactly because they simulate a fish's eye.

b) **Landscapes**: it uses landscaped to represent the interrelationships of data groups containing related metaphoric images [18]. A tool which uses this technique is the *kartoo* (www.kartoo.com), a web metasearch site.

c) **Parallel Coordinates**: it was proposed by Inselberg [10] and allows relating pieces of information among each other. The technique emphasizes relationships between adjacent axes and sets of data which have similar standards.
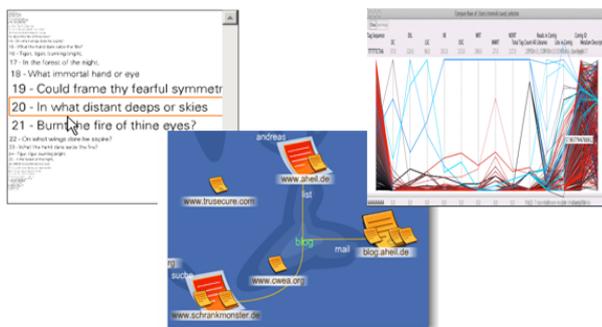
Figure 1 - Information Visualization Techniques.

2.2   Visual Data Mining

Visual data mining consists in the integration of Data Mining techniques and Information Visualization techniques. With the help of visualization techniques, users can discover new and useful knowledge. The goal is to identify information implicit in great volumes of data. To be effective, Data Mining process has to include the user for exploring data and matching flexibility, creativity and domain knowledge. The basic idea of visual data mining is to present the data in some visual form, allowing the user to obtain a view of the data, making up conclusions and interacting directly with the data [11].

2.3   Related Works

Many information visualization tools have been developed to facilitate human understanding, such as *LifeLines*, *Chat Circles* and *TimeMines*. *LifeLines* [14] is a novel technique that generates a visual representation about patients' medical life, including admissions, diagnoses, medications and improvements. *Chat Circles* [20] is a graphic interface for simultaneous conversations, which attempts to reveal activities patterns in interactions among users. And *TimeMines* [19] it is a tool which generates a visual representation with words and its relationships in periods of time analyzed.

VisMade tool, presented in this paper, is different from these and other tools, because it organizes data in the master-detail format.

3   VisMade Visualization Tool

Many organizations have data structured into the "*master/detail*" format, or *one-to-many* relationship. In this format, there is a main concept and a table with relationships. The visualization tool developed here is available for use in the electronic address www.edecio.com.br/visinfo.

The idea of developing this tool arose from the observation that many applications such as stock control, academic control, medical offices, hospitals, etc., have a similar structure with master-information and detailed-information (products and product sales; patients and their consultations over a period of time). For each "master" table, there is a "detail" table, with indications of occurrences in the "*master*" table.

Observe the typical models of structuring these tables in figure 2, which presents the relationship between the Movie tables (master), and Rentals (detail) of a traditional movie rental control system of a video shop. In this case, each movie may be rented many times and each rental has information such as customer identification, data and value.
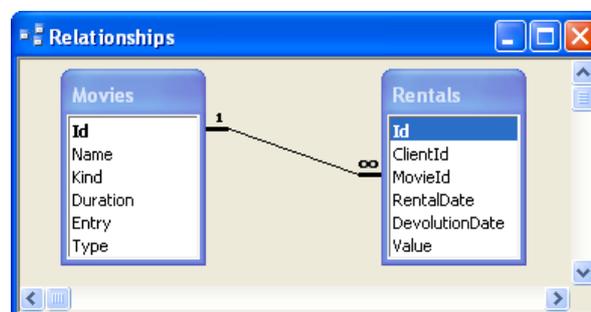
Figure 2 - Table structure in the master/detail (one-to-many) model

This format explores one-to-many relations such as product-sales, patient-visits, disease-patients and so on. Similarly to snowflake and star schemas from OLAP

technology (or data warehousing), the master-detail schema also deals with multidimensional data (for details on On-Line Analytical Processing, see Chaudury and Dayal, 1997 [6]). There are functionalities such as *rollup* (increasing the level of aggregation) and *drill-down* (decreasing the level of aggregation or increasing detail) along one or more dimension hierarchies, *slice_and_dice* (selection and projection), and *pivot* (re-orienting the multidimensional view of data). The difference is on that VisMade represents data in a graphical format, allowing quick analysis and thus rapid discovery of new information. Besides that, VisMade is more appropriated to deal with temporal data, showing time lines with different colors and thus helping in more quick analyses.

3.1 Data Types

One of the most important features of the proposed tool is the fact that it is adaptable to the data submitted by the user. For the correct working of the tool the data must have a structure like master-details, described as follows:

Master Table
- must have, at least, one attribute (field) to be used for data list exhibition, from which the data will be grouped and on which filters may also be defined;
- must have one attribute which sets the relationship with the detail table, being them of the same type;

Detail Table
- must have one attribute which creates the relationship with the master table (foreign key);
- must have an attribute that indicates the event occurrence date to be analyzed, which will be exhibited in the temporal axe of the visual representation;
- must have one or more attributes of numeric type on which the calculi to be represented in the tool will be performed. These attributes are recommended, but not mandatory, since the basic calculus of the interaction occurrence counting may be exhibited, from the two attributes indicated above.

3.2 Development Methodology

In order to build a new Information Visualization tool, according to Ferreira and Nascimento [8], a methodology comprising the following steps should be used: a) study of the problem and of the data to be visualized; b) visualization building; c) definition of interaction mechanisms; d) a prototype system implementation; e) visualization evaluation.

The paradigm of visual exploration proposed by Keim [11] was also used as reference for the construction of the tool, in which the data visual exploration must follow a three-stage process: general view, zoom & filter and detail per demand.

The tool evaluation process is being done by means of two sets of evaluation criteria defined by Luzzardi [13], to evaluate Information Visualization techniques: one related to the visual representation and other related to the interactive mechanisms which are present in the tool.

3.3 Technique Used by the Tool

The most suitable visualization technique for representing data structured in master/detail (one-to-many) format is Parallel Coordinates, where groups of registers from the master table are represented in timelines, while the values obtained in the detail table are represented in "Parallel Coordinates" of the technique. Observe the visual representation exhibited in figure 3, which shows movie rentals grouped by gender. With the Parallel Coordinate technique it is possible to:
- use the initial column to exhibit the groups from the main table;
- make each of the visual representation columns have a period to be analyzed (in the example, the months of the year);
- demonstrate the approximate group rental values in each of the months through the lines passing on these columns;
- exhibit after the last column the strip and represented values.

Figure 3 - Example of Parallel Coordinate Technique Use

When filling out the form with the indication of the master table and the details table, the tool generates an initial visualization with which the user may interact. Clicking on the items exhibited in the graphic, organized according to the hierarchy defined by master and detail tables, it is possible to the user to amplify or restrict the visualization of the subitem of each clicked item (figure 4)
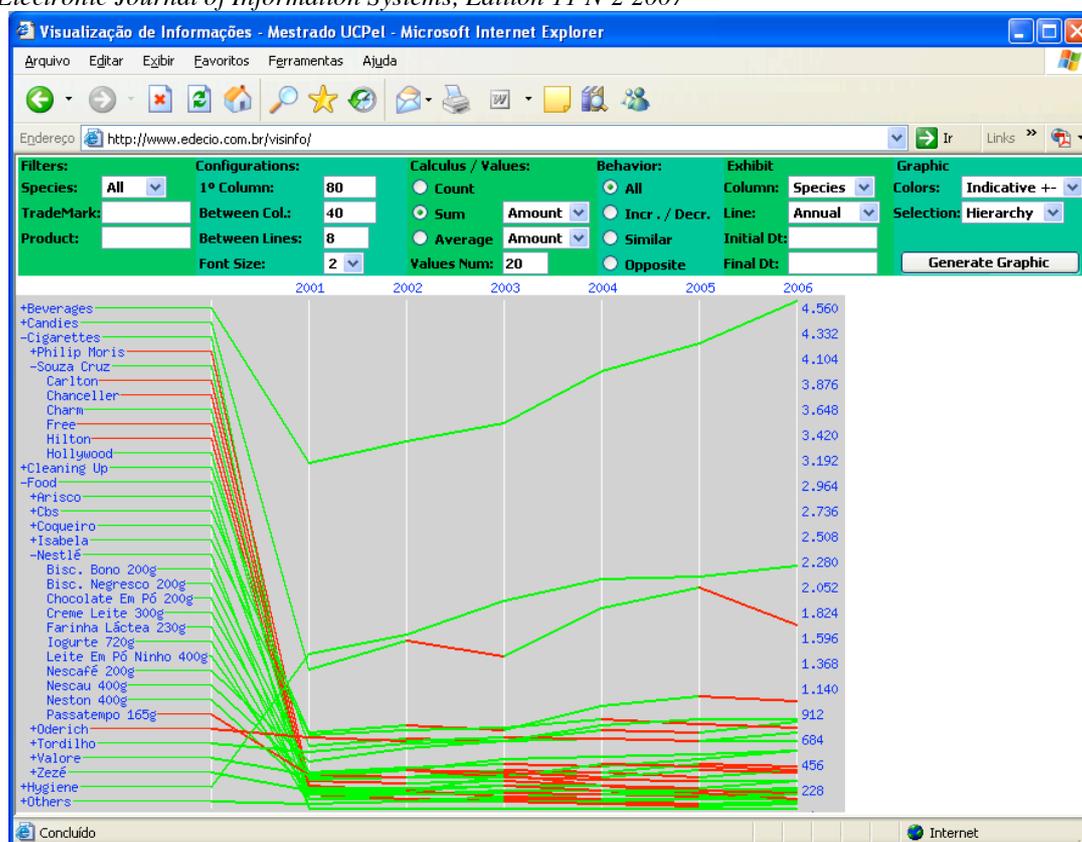
Figure 4 - Use of the tool in a traditional sales control system

Among the various options of interaction available in the tool we can point out:

Colors: allow the user to indicate the color to be used in the graphic. He/she can opt to use different colors for each item presented or colors that indicate increasing (+) or decreasing (-) in the interactions of the detail table. The different colors allow following the performance of an item along the visualized period. The indicating colors serve to give the user a general view of item performance from one period to the other. For example, if from 2002 to 2003 the green color is prevalent over the red one, this is a sign of general improvement on services in this period.

Selection: serves to define if the selection of an item in the graphic will make the tool amplify or reduce the detailing level of the presented items or point out this item in comparison with others. However, if Point-Out is selected, clicking on an item modifies its color, making it different from the others. The user may point out/match as many items as he/she wishes, each of them being exhibited in a different color, according to illustration in Figure 5.

Behavior: makes it possible for the user to point out the items which present increasing or decreasing behavior, similar behavior, or which behave in an opposing manner in the analyzed period. In the case of

items with opposite behavior, a pair of light/dark colors indicates the relationship. In the case that the number of lines to be pointed out is superior to ten, the tool makes the pointing out paging, avoiding that any relationship may be unnoticed by the user. Figure 6 shows the items with similar behavior, while figure 7 points out the items with opposite behavior. In the example shown in the figure 7, it is possible to see that "cigarettes" and "candies" have opposite behavior leading to the hypothesis that they are competing products.

Calculus: indicates the type of calculus to be applied on the items of the detail table. The user may select among Adding, Counting or Calculating the average value. In the case of adding or average it is also necessary to indicate the field on which the calculus will be performed.

Several other tool configuration options may be modified by the user, like the space between the lines, the source size, the time interval to be exhibited in the graphic, etc. Besides that, **VisMade** executes directly from the web, making it possible from data exporting via XML to the tool until the generation and interaction of the user with the generated visual representation.
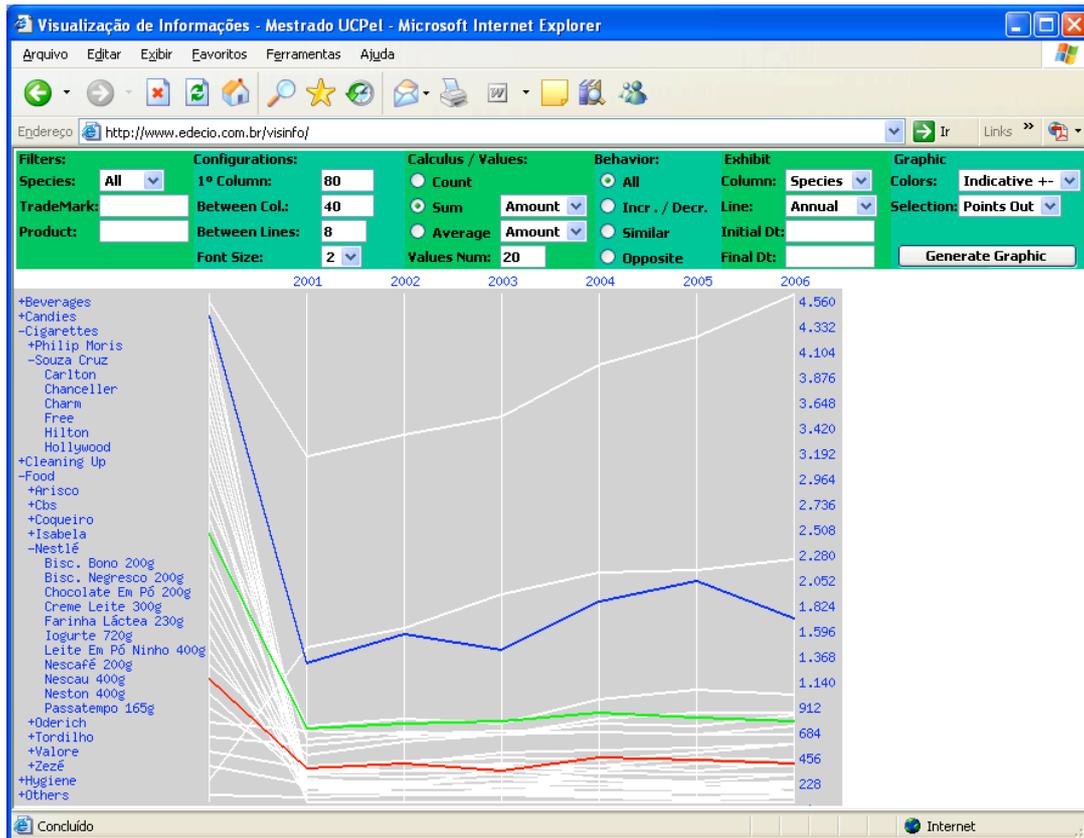
Figure 5 - Selecting items for comparing performances
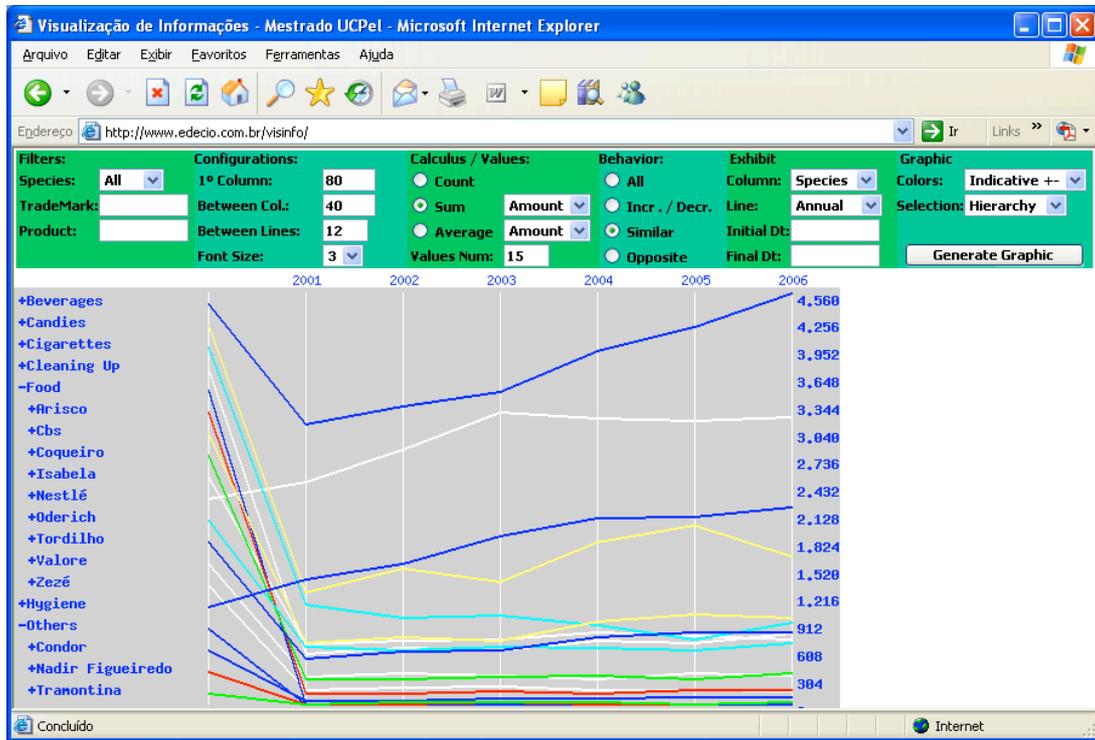
Figure 6 - Items with similar behavior
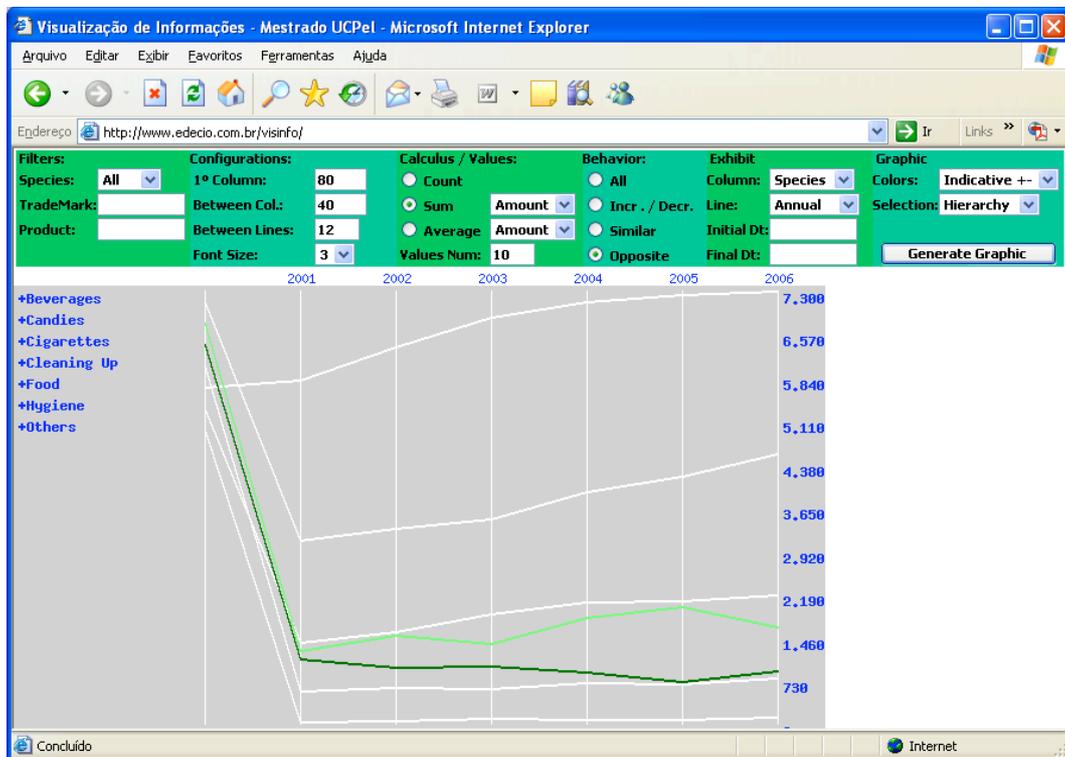
Figure 7 - Items with opposite behavior

4   Case Study

For better understanding the characteristics of the presented visualization tool, an example of application of the tool will be presented and discussed. The application refers to real data from a Medical Clinic. There is a database with information about patients, their demographic data, their admissions in the Clinic and relationships to diseases. There are also data about physical places in the Clinic (wing and rooms). Diseases have symptoms associated to them, following a guide created by experts.

The VisMade tool allows users to view data under different perspectives, according to the multidimensional criterion from OLAP technologies. Following, we discuss the different ways to view those data.

a) **Wing, Diseases, Admitted Patients:**

This structure allows a general view about the number of admissions in each wing of the Clinic. Information about diseases is important to map the kind of admissions per wing along the time and discover wings where patients with non similar diseases are allocated. This kind of event is not useful to the Clinic, since physicians have to move from one wing to another in order to attend patients. The sum of "days of admission" (period of admission) allows inferring which diseases need more time and which ones demand short periods; one of the goals of the Clinic is to shorten the admission period. Analyzing opposite behaviors allows discovering which wings or diseases may be complementary for admissions, that is, in which time periods one wing is too occupied and other wing has vacancies.

b) **Symptoms, Diseases, Admitted Patients:**

This point-of-view is useful to analyze which symptoms generate more admissions in the Clinic. Yet it is possible to discover in which time periods during the whole year certain symptoms are more frequent. Under this view, it is also possible to discover whether there are symptoms with similar or opposite behavior (frequency).

c) **Ages, Diseases:**

When there are numerical fields, the tool divides instances in 10 groups. In the case of ages, each group represents an interval of ages. This view allows identifying relations between age intervals and diseases along the year (Figure 8). Also we can analyze seasonal behaviors for each age group, discovering periods where there are more admissions for each age interval.

d) **Diseases, Patients Attributes:**

Relating diseases with some patient's attributes (such as age, city, economic activity), it is possible to discover the diseases that are more associated to some details and vice-versa. Besides that, we can analyze whether there are periods where diseases increase the number of admission cases into the Clinic.
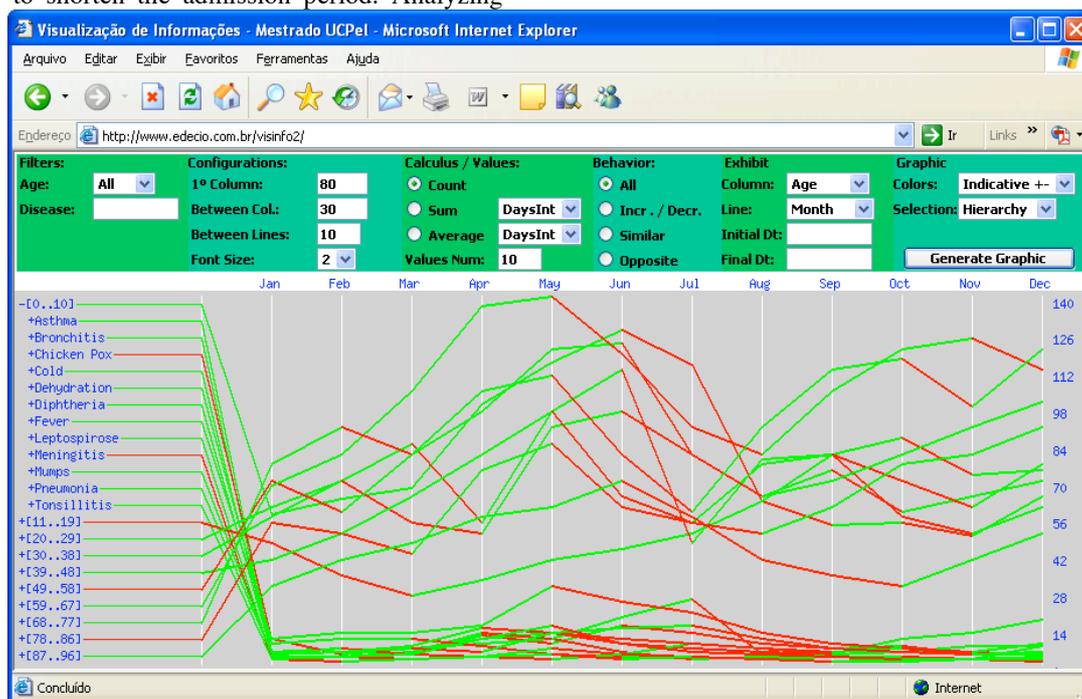
Figure 8 - Data from a Medical Clinic under perspective of Ages and Diseases

5 Conclusions and future works

This paper described a visualization tool created for the analysis of information structured in master-details formats and for the discovery of new knowledge (visual data mining). Besides that, the tool is appropriated for visualizing data during time periods. Its functionalities allows the analyze of items with similar or opposite behaviors and quickly the discovery of trends in the data by showing increasing and decreasing behaviors.

A case study of a Medical Clinic was presented to demonstrate that a database may be viewed by different perspectives, following the master-details structure. The case study also helps in demonstrating what kind of discoveries may be performed with the tool.

The next step is the application of the tool in different cases for comparing performance and for analyzing which kinds of applications may benefit from the use of a visualization tool such as **VisMade**.

6 Bibliography

[1] AGRAWAL, R. "Data mining: the quest perspective" In: EDBT SUMMER SCHOOL, Gubbio, Italy, 1995.

[2] AGRAWAL, R.; IMIELINSKI, T. "Database mining: a performance perspective" IEEE Transactions on Knowledge and Data Engineering, New York, v. 5, n. 6, 1993

[3] AHLBERG, C.; SHNEIDERMAN, B. "Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays", In: ACM Conference on Human Factors in Computing Systems, pages 313–317, 479–480, 1994

[4] CARD, S.K.; MACKINLAY, J.D.; SHNEIDERMAN, B. "Information Visualization", In: Readings in Information Visualization - Using Visualization to Think. San Francisco, Morgan Kaufmann Publications, 1999

[5] CAVA, R. A.; CHUBACHI, O. M.; FREITAS, C. M.; LUZZARDI, P. R. G. "Introdução à Visualização de Informações". UFRGS: Revista RITA, Porto Alegre, 2001

[6] CHAUDURY, S.; DAYAL, U. "An overview of data warehousing and OLAP technology". ACM SIGMOD Record, vol. 26, no. 1, pp. 65-74. 1997

[7] FAYYAD, U.; et al. "Advances in Knowledge Discovery and Data Mining." Cambridge: MIT Press, 1996

[8] FERREIRA, C. B. R.; NASCIMENTO, H. A. D. "Visualização de Informações – Uma Abordagem Prática". Congresso SBC, São Leopoldo, RS, Brasil, 2005

[9] GERSHON, N.; EICK, S. "Information Visualization. IEEE Computer Graphics and Applications", Los Alamitos, v. 17, n. 4, p. 29-31, July/Aug, 1997

[10] INSELBERG, A. "Multidimensional detective. In IEEE Symposium on Information Visualization", pages 100–107, Washington, Brussels, Tokyo. October, 1997

[11] KEIM, D. A. "Information Visualization and Visual Data Mining". IEEE Transactions On Visualization And Computer Graphics, v. 7, n. 1, January-March, 2002

[12] LOH, S.; GARIN, R. S. "Web Intelligence – Inteligência Artificial para Descoberta de Conhecimento na Web". http://www.inf.unisinos.br, 2001

[13] LUZZARDI, P. R. G. "Critérios de Avaliação de Técnicas de Visualização de Informações Hierárquicas". Tese de Doutorado. UFRGS: Porto Alegre, 2003

[14] PLAISANT, C.; SHNEIDERMAN, B. "LifeLines: Using Visualization to Enhance Navigation and Analysis of Patient Records". Revised version appeared in 1998 American Medical Informatics Association Annual Fall Symposium, 1998

[15] RAO, R.; CARD, S. K. "The table lens: Merging graphical and symbolic representation in an interactive focus + context visualization for tabular information" In Proceedings of ACM Conference on Human Factors in Software, ACM-Press. 1994

[16] ROBERTSON, G.; CARD, S.; MACKINLAY, J. "Cone Trees: Animated 3D Visualizations of Hierarchical Information" Proceedings of ACM CHI´99 p.189-194. 1994

[17] SHNEIDERMAN, B. "The eyes have it: a task by data type taxonomy for information visualizations" Proceedings of IEEE Symposium on Visual Languages, Boulder, CO, September 3-6. p. 336-343. 1996

[18] SPENCE, R. "Information Visualization". ACM Press and Addison-Wesley, 2001

[19] SWAN, R.; JENSEN, D. "TimeMines: Constructing Timelines with Statistical Models of Word Usage" Department of Computer Science, University of Massachusetts - Amherst, Massachusetts USA, 2000

[20] VIEGAS, F.; DONATH, J. "Chat Circles" Computing System, New York: NY, USA. ACM Press new York, NY, USA, 1999

[21] WARE, C. "Information Visualization Perception for Design", San Francisco: CA, USA. Morgan Kaufmann Publishers, 2004

Authors

Edécio Fernando Iepsen is a Mastering student from the Computer Master Degree Course at Catholic University in Pelotas. He has experience on Systems Analysis and Development, whose created software are been used on the market. Nowadays, he is a Professor at Senac College, teaching disciplines like algorithms, language programming, e-commerce and systems security.

***

Paulo Roberto Gomes Luzzardi is a Professor in the Catholic University of Pelotas, in Brazil. He has a PhD degree in Computer Science, obtained in 2003 at the Federal University of Rio Grande do Sul. His research interests include Computer Graphics, Information Visualization, Digital TV and Sign Language (LIBRAS and Signwriting). He acts as reviewer for IASTED - International Association of Science and Technology for Development and others.

***

Stanley Loh is a Professor in the Catholic University of Pelotas and in the Lutheran University of Brazil, in Brazil. He has a PhD degree in Computer Science, obtained in 2001 at the Federal University of Rio Grande do Sul. His research interests include recommender systems, data-text-web mining and technology applied to knowledge management and business intelligence. He acts as reviewer for periodicals like IEEE Transactions on Knowledge and Data

Stanley Loh is a Professor in the Catholic University of Pelotas and in the Lutheran University of Brazil, in Brazil. He has a PhD degree in Computer Science, obtained in 2001 at the Federal University of Rio Grande do Sul. His research interests include recommender systems, data-text-web mining and technology applied to knowledge management and business intelligence. He acts as reviewer for periodicals like IEEE Transactions on Knowledge and Data Engineering, IEEE Internet Computing and others.