

TIME HISTOGRAMS WITH INTERACTIVE SELECTION OF TIME UNIT AND DIMENSION

Snezana Savoska¹, Suzana Loskovska², Vlatko Blazeski¹

¹Faculty of Administration and Management of Information systems, University St. Kliment Ohridski
Bitola, Partizanska bb, 7000 Bitola
e-mail: savoskasnezana@yahoo.com

²Faculty of Electrical Engineering and Information Technologies, , University Ss. Cyril and Methodious
Skopje, Karpos 2 bb, p.f.574, 1000 Skopje, Republic of Macedonia

ABSTRACT

Many researchers are working on the improvement of the usability and efficiency of the sale data analysis methods which are required by the users – analytical staff and managers. In this paper, we present the improved technique of time histograms. We named proposed method Time Histograms with Interactive Selection of Time Unit and Dimension (THISTUD). Performed modifications and developed interactive user interface are described. The system is tested on the data warehouse that includes data for the ten years. The created sale data mart and visualization results are presented in the paper too.

1 INTRODUCTION

The use of visualization in the business environment as a method for an efficient information gaining is increasing rapidly, especially in the processes where the quick response time is required like those as data analysis and decision making [9]. The common methods for business information visualization are the x-y diagrams [3], histograms [1] and the combination of scatter-plots [5]. When the histograms are used for business information visualization, the data are commonly displayed as two or three dimensional shapes [12]. Even the sale quantity and the amount of income are usually selected as the histogram axes, time is often required as additional dimension, too. In 2D histograms, the visibility of the sale quantity and income is not satisfactory if the time is not included as an additional attribute. The visibility is very important when the multidimensional data like data from data warehouse is analyzed and correlations between variables are explored, for example the season trends. In that context, 2D histograms are improper for visualization because the user can see only one dimension at the same time on the same display. Therefore, we developed a method to improve the visualization introducing possibility for interactive selection of the time unit in the main window and an additional window to visualize data for the selected time unit. The main window provides general overview of the multidimensional time-dependent data. In the additional window, the data for the selected time unit is shown. Interaction is provided for the second window, too, enabling

users to select data that is visualized in a new sheet (tab) of the window. The number of tabs is not limited. We used only two tabs in our example for visualization of sale regions or subregions for ten years.

2 METHOD THISTUD

Information visualization in 2D produces good results for analysis when using the histogram bars and x-y diagrams. The users can analyze data with drill-down property through dimension. This user-friendly feature is gained using OLAP Analyses services and OLAP cubes and their visualization possibilities [13]. Figure 1 shows the OLAP pivot chart for five years sale data. Data is presented by 2D histogram bars with different colors. The Pivot table lists all dimensions and measures. This tool offers interactive tools for data filtering enabling users to select one or combination of time period, city of sale and article. Results are presented according to user's requirements.

But, when the problem is multidimensional, the 2D techniques are limited and can not satisfy the user's requirements. Managers usually face this problem, when analyze the sale data for several years, depending on sale region, sub-regions or period of analysis.

Therefore, we developed the interactive method for multidimensional information visualization based on histograms. Instead of using one visualization plane, we use two windows. The main window displays general overview of data. The user can select time unit from the main window by added interactive tool and the data for selected time unit is represented in the second window. Interaction is added to the second window too, allowing users to change dimensions of data visualized in this window. The change of the dimension is possible by tab menu. We named proposed method Time Histograms with Interactive Selection of Time Unit and Dimension (THISTUD). The interface of developed system based on this method is presented in the Figure 2 and Figure 3.

In our previous paper, we proposed the method Time Histograms With Select and Zoom (THWS&Z) [6] with interactive possibilities but this method had a good response time for problems with a small number of values for each

dimension. When the dimension values are increasing, the response time do not satisfy the user's requirements. Performed improvements of the THWS&Z were embedded in the method THISTHUD. The THISTHUD consists of the following steps:

- ❖ Preparing data in the aggregated tables
- ❖ Visualizing data in the main window
- ❖ Determining the slighter and selection visibility
- ❖ Waiting time unit selection
- ❖ Visualization of data for selected time unit

- ❖ Enabling selection of dimension
 - ❖ Set up of ToolTip property in the second window
- Data repositories with source data for visualization are created by the following procedure:
- ❖ Loading data from Transactional databases
 - ❖ Data cleaning and loading in Data Warehouse or data marts [2]
 - ❖ Creating the additional views and aggregated tables if they are necessary

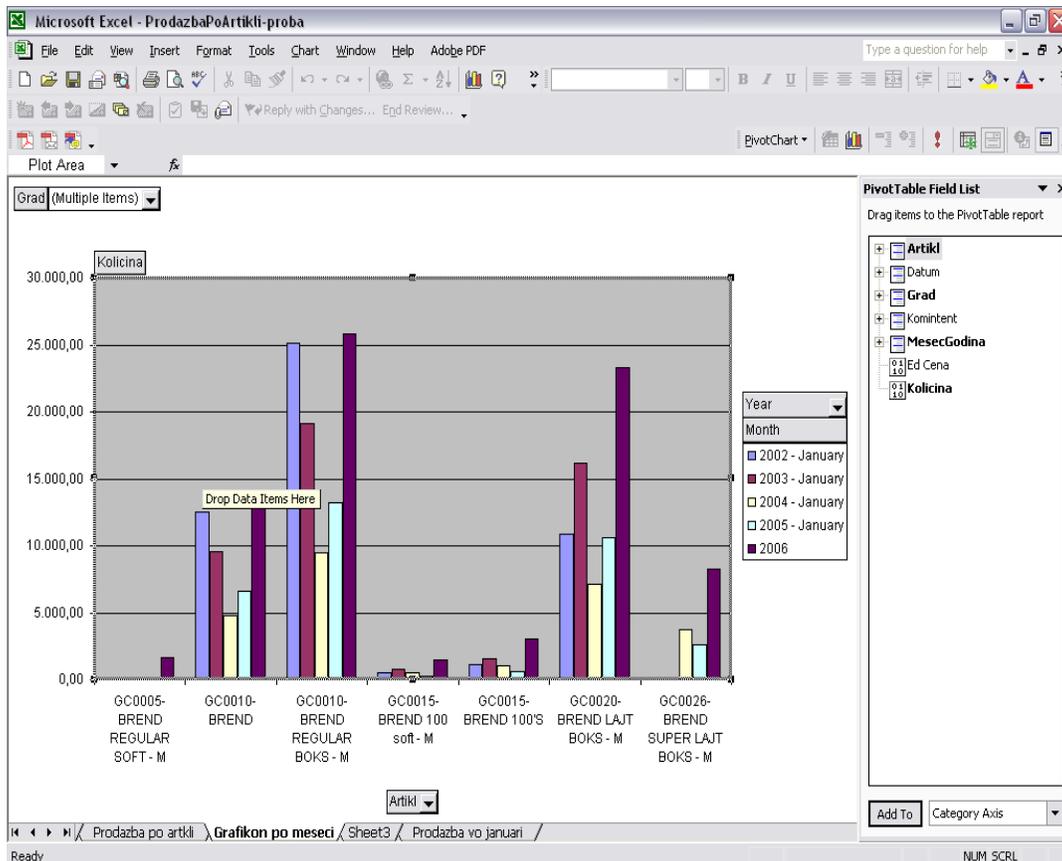


Figure 1: OLAP visualization wizard

3 THE PROBLEM OF SALE DATA VISUALIZATION

Our method is especially useful for visualization of monthly, quarterly or yearly sale quantity, financial data or income analysis, because it provides a global view of data and enables analysis of periodical data. We tested our system to the problem of visualization of sale data.

The sale data visualization requires possibility to change sale period and analyze data for many dimensions. We obtained the sale quantity data for ten years for Production and Trade Company, prepared for monthly analysis. During preparation phase, we created the aggregated tables in data mart with unique dimension codes [2]. For this purpose, the additional codes are introduced and data is

grouped according the dimensions (data as region, sub-region, city, production group or time slab in which sales are made). This approach by using Data Warehouse or data mart provides advantages because data can be updated each time when new data is added to the original databases [6]. The time period of 10 years is cited in month periods – time slabs. These time intervals represent the general dimension for this purpose [6]. Because the most important things for sale data analysis are the sale quantity and sale value, they will be analyzed for many independent variables – dimensions – regions, sub-regions, cities, production groups and other desired dimensions.



Figure 2: Time histograms with interactive selection of time unit and dimension



Figure 3: Time histograms with interactive selection of time unit and dimension

4 RESULTS

Figure 2 and Figure 3 present visualization for the sale data. The main window contains the visualization of the total sale data for time slabs. The user can see the sale trends over the years/months of all products and regions. The period of one month is selected as a unique time slab, but the period can be smaller as a half month, a week, a day or bigger like a quarter, a half-year or an entire year. The value of the selected time slab (x-axis) to which the pointer points is shown in the right upper corner. The pointer is a vertical line

with color that is different from the bars color. The pointer enables users to select the time slab which data will be presented in the second window. The analytical values are shown on the second window. The system enables each time slab to be visualized in separate sheet in the lower window.

The user can move between sheets in the additional window by tab control button. When the region is selected by tab control, the user can analyze sale data for selected time slab for sale regions. Each region is presented with different color to provide easy recognition of the visual

representation. The value of sale are shown over each region bar because the ToolTip property is set.

The dimension in the additional window can be changed to represent region or sub-region (or city, product group,) for selected time slab. In the case presented in the Figure 3, there are ten different bars with values of the displayed measures for sub-regions, corresponding to the existing ten sub-regions. If the cities are required as dimension for analysis, there will be 35 bars with assigned values and corresponding legend for each time slab (35 cities for sale analysis). In this manner, in the same windows, the user can analyze many dimensions of the problem.

5 FUTURE WORK

There are many possibilities for further improvement of this visualization technique. These improvements can be made in data acquisition, data preparing and visual data rendering phase and according to users needs and feedback of using this technique.

Data acquisition improvement can be done by adding trigger procedures that allow automation of data refresh in data repository. Using transitional databases helps in this case.

In the process of interaction programming, it is possible to define criteria for data selection and data filtering. For example the user can define the lower and upper sale border line. The radio buttons can be added to enable data selection and filtering criteria – all data in the graph, the sale data below the lower border line of sale, or data between the two border lines of sale.

Visual data delivery can be improved by creation of aggregated tables when the process of data updating in the Data Warehouse, located in the server site, is done. This method is known as batch data processing. In this way, the data is prepared for faster visualization and the response time is much shorter, compared to the response time in on-line data visualization. This is one of the less desired methods for data processing but useful for improving desktop data delivery response time.

6 CONCLUSION

The method Time histograms with interactive selection of time unit and dimension (THISTUD) enables visualization of multidimensional data. Providing two windows for data visualization, the user can interact with the data on-line. The system provides double interactivity tool enabling user to select time slab of the main window and to change the view of data in second window. It is obvious that this visualization tool provides good analytical possibilities and

a solid base for visual pattern and trend recognition, especially when the selecton and filtering options are set.

The possibilities for further adjustment of this method to provide visualisation of different problems are practically unlimited. The user only have to define their requirements – topics of analysis, data filtering, knowledge-based patterns or criteria introducing for selection, clustering or some kind of data mining.

References

- [1] Kosara R., Bendix F., Hauser H., TimeHistograms for Large, Time-Dependant Data, VrVis Research Center, Vienna, Austrija, IEEE TCVG Symposium on Visualization (2004)
- [2] Bonifati A., Cattaneo F., Ceri S., Fugetts A., Paraboschi S., Designing Data Marts for Data Warehouse, ACM No.4, 10.2001, pp. 452-483
- [3] Hansen C., Johnson C., The Visualization handbook, USA-2005, ISBN 0-12-387582-X
- [4] Ware C. Information Visualization, Perception for design, ISBN 1-55860-819-2, 2004
- [5] Wong P.C., Bergeron D.R., 30 year of Multidimensional Multivariant Visualization, ACM, grant NSF IRI-911753
- [6] Savoska S., Loskovska S. Blazeski V. Time Histograms with select and zoom, ICEST 2008, Nis, Serbia, pp. 445-448
- [7] Card S., Grudin J., Nielsen J., Information vidualization – Perception for design – Second edition, 2004
- [8] S.O.Tegran, T.Keller (EDS), Knowledge and Information Visualization, - Information and Knowledge visualization in development ad use of a MIS for DC, 2005
- [9] Zhing J., Croft W.B., The Information Retrieval Series, ACM Code: H.3,H.4,H.5, pp 17-23
- [10] Thaper N., Indyk P., Guha S., Koudas N., Dynaic Multidimensional Histograms, ACM SYGMOD' 2002, June 4-6,
- [11] Harve S., Heitzlr B., Nowell L., ThemeRiver: In Search of Trends, Patterns and Relationship, ACM
- [12] Akiba H., Ma KL, A Tri-Space Visualization Iterface for Visualizing Time-Varying Multivariate Volume Data, Eurographics/IEEE-VGTC-Symposium on Visualization (2007)
- [13] Designing and Implementing OLAP Solutions Using Microsoft SQL Server 2000, Microsoft Course 2074