



“Mapping Techniques and Visualization of Statistical Indicators”

Haitham Zeidan

Head Of electronic Dissemination Division/ Information

System Directorate

Palestinian Central Bureau of Statistics(PCBS)

Ramallah – Palestine

Tel: 00 (972/970) 2-298 2700

Fax: 00 (972/970) 2-298 2710

E-Mail: Haitham@pcbs.gov.ps

Abstract

Statistical data has gained many interests from policy makers, city planners, researchers and ordinary citizens as well. Illustrating data in a communicative format helps decision makers to understand and analyze effectively large amount of information in a short time. data visualization, as a way of presenting different data types in a more understandable form, is growing increasingly in various areas. There have appeared numerous statistical data visualizations; Unfortunately, even though there has appeared a great deal of statistical data visualization, applications for visualizing statistical data are still rare. Moreover, these seldom applications also suffer drawbacks. This paper aims to investigate how data visualization could be used to increase readability and usability of statistical data. different visualization techniques and interaction techniques were used based on Highcharts Java Script libraries [1]. In order to identify user needs interview with expert



users have been done, second interview have been done also to evaluate our final visualization application, the evaluation showed a high level for both effectiveness and efficiency of the visualization techniques in our application. Moreover the majority of the evaluation participants were satisfied with the application, because it has helped them to accomplish all tasks successfully.

Key words

Data Visualization, Indicators, Ontology, PCBS, Web 2.0, Data Integration, Interactive techniques.

1. Introduction

Official statistics are statistics published by government agencies or other public bodies such as international organizations. These statistics provide quantitative and qualitative information on all major areas of citizens' lives, such as economic and social development, living conditions, health, education and the environment. Official statistics can be found on web sites of national statistical agencies such as the Palestinian central bureau of statistics (PCBS) [2]. There is a need for a common schema to integrate that massive data and visualize the findings, so that viewers can easily derive an insight into data. The objective of the integration is to perform the data harmonization acquired from different sources and files, mapping algorithms could be used to map indicators [3].

The major tenets of Web 2.0 are collaboration and sharing, The term 'Web 2.0' has become undisputed linked with developments



such as blogs, wikis, social networking and collaborative software development. Web 2.0 can make dramatic impact on developing interactive and collaborative visual analytics tools for the Internet. Tools are needed that advances humans ability to exchange gained knowledge and develop a shared understanding with other people [4].

2. Objectives

Due to the increasing complexity and heterogeneity of statistical data, an increasing need for sophisticated visualization technology arises, We introduced a new visualization application to enhance presentation of official statistics based on visual analytical approach that combines both data analysis and interactive visualization. We implemented in our previous work new mapping algorithm [3] based on hamming distance [5], edit distance [6] and ontology, using our algorithm we enhanced integration, harmonization and mapping of statistical data indicators from different sources, the indicators after importing saved as Resource Description Framework (RDF) in the schema, this will help to visualize statistical data using visualization techniques, the schema were built included ontology tables to improve and increase the accuracy of the mapping algorithm. We tested the accuracy of the algorithm, experimental results shown high accuracy of mapping for the algorithm by adding the ontology to the algorithm. In this paper we extended our work to enhance presentation of statistics based on visualization and interaction



techniques that make humans capable of analyzing data by presenting results in a meaningful and intuitive way while allowing to interact with the data. Visualizing statistical data help promote the use of statistical data for improved planning and policy making. all visualizations were implemented with the help of Highcharts Java Script libraries [1].

3. Application Implementation

The visualization in our application were implemented based on the Microsoft platform and .NET Framework using the software development tools of Microsoft visual studio, including .NET and ASP.NET., Highcharts Java Script libraries [1], Highcharts is a charting library written in pure JavaScript, offering an easy way of adding interactive charts to web site or web application. Highcharts currently supports line, spline, area, areaspline, column, bar, pie, scatter, angular gauges, arearange, areasplinerange, column range and polar chart types. Searching, comparing, re-visualizing, sorting, and filtering interaction techniques were implemented for visualizations in our application, The purpose of our application to store data and indicators, for organizing, storing and presenting data in a uniform way, and present information in the form of tables, graphs and maps, and to facilitate data sharing. the application supports several tasks to do visualization.

4. Evaluation



This section describes the approach used to specify users requirements and to evaluate the visualization application. The first step is done by making interview to get requirements before implementation phase of visualization techniques for our application from the end users. The second step is realized by making interview to get and find out the requirements and feedback of end-users for our final application and visualization techniques that we used. Further sections describe the process of evaluation in greater detail.

4.1 Interview with expert users to specify users requirements

This step is done by making interview to get requirements before implementation phase of our visualization application from the end users, We collected feedbacks from interview with 13 participants. Participants were selected to be representative of the intended user community, including Statisticians, researchers, decision makers so that they could specify user requirements related to visualization, design appropriate functionalities, and develop a visualization application for statistical data, we tried to specify user requirements related to visualization, design appropriate functionalities, to build visualization system according to the users requirements. Our findings from this interview are concluded as below:

- All the users reported that they interested in visualizations, most of the users interested in line charts, pie chart, bar chart, column chart, and map chart. nine of the users mentioned that



the line chart a more convenient way to explore the results than other techniques.

- both expert and end users are interested in a visualization statistical data and they think it could facilitate their work and can help them to have a better and clearer understanding of the results.
- They suggested to use different types of graphs and charts describing the statistical data and comparing different time series, visualize system should support different languages, using maps and colors, adding features to download and share the visualization output and they suggested also to use animation for time series and using color gradient, to be more flexible when sorting according to time series data and going through items and sub items.
- 11 users interested in comparing different scenarios of statistical data this could be an important feature of the visualization model. they think these features can help them in evaluating the output results and facilitate the decision making process.

4.2 Interview with end users to evaluate the visualization and interaction techniques in the application

A structured interview with pre-defined questions performed. 12 participants were asked to use the visualization techniques of the application to perform a number of preselected tasks that are related to participants work context. Participants were asked to



think aloud while carrying out the tasks. At the end participants were asked to fill in a feedback questionnaire on the overall use of the application and estimate their answers from 1-"absolutely disagree" to 5-"absolutely agree". All questions were carefully selected in order to make conclusions about effectiveness, efficiency of the application, and the level of the user satisfaction. All of those people are Statisticians, researchers, decision makers so that they could really estimate whether the new visual application capability are effective and efficient enough for solving certain tasks from that research field. In the feedback questionnaire there were also open questions, where the participants of the evaluation could write their opinions and comments. The answers on those questions are very important. Summarizing the results of the interview done with the help of feedback questionnaire, the following conclusions about effectiveness, efficiency, and user satisfaction of the application:

- Almost 83% of the participants were absolutely agree, that in order to solve the tasks the application was easy to use. Also those people stated that they felt confident about the results they have got after some tasks were accomplished.
- Almost 67% of the participants absolutely agree and liked the integration of the different views of the structure of the data.
- almost 82% of the participants of the evaluation were disagree that they should spend much time to complete the tasks or that they were often confused during the tasks accomplishment.

Also there were some requests for additional functionality:



- It would be great if the application can do some basic statistical analysis.
- It will be very useful if there is an option to add more than one indicator in the same chart (where its applicable and could be executed), like when we need to show more than one variable in the same line chart (like if we need to show the values through time series for Number of household and the average household size at the same year, or the population and the number of unemployed individuals).
- Supported of different languages.
- Animation and color gradient.
- More tooltips in order to explain the meaning of the indicators.

5. Conclusion and Future Work

This research aimed to introduce a new visualization application for presenting statistical data, We enhanced presentation of statistical data based on dynamic visual user interfaces and the principles of Visual Analytics. this application has been introduced to provide techniques that make humans capable of analyzing data by presenting results in a meaningful and intuitive way while allowing to interact with the data. Visualizing statistical data help promote the use of statistical data for improved planning and policy making. all visualizations were implemented with the help of Highcharts Java Script libraries, which has helped to make highly interactive and hence responsive for the end user visualizations.



The application was successfully evaluated by different users and experts, statisticians, researchers, decision makers. the evaluation results have shown high levels for effectiveness, efficiency, and for the user satisfaction of the framework. future work includes improving collaboration the application. Additional methods are required to support the users in finding good views on the data and in determining appropriate visualization techniques. we have to consider the 3D visualization of uncertain graph structures with uncertain attributes, which we think is a formidable challenge.



Acknowledgements

I'd like to acknowledge PCBS to give me this chance to attend a conference like IAOS.

I am extremely grateful and indebted to my colleagues in PCBS, for helping me in preparing the statistical data.

My thanks are extended to all those people, who have participated in the evaluation process. Thank you for taking time for it.

I thank my mother for her love and support, my father for the encouragement and motivation, and other members of my family for being there.

Finally, and most importantly, I would like to thank my wife. Her support, encouragement, quiet patience and unwavering love were undeniable.



7. References

[1] Highcharts library written in pure JavaScript:
<http://www.highcharts.com>.

[2] Palestinian Central Bureau of Statistics (PCBS):
<http://www.pcbs.gov.ps>.

[3] H. Zeidan, R. Jayousi and J. Najjar: Interoperable Visualization Framework Towards Enhancing Mapping and Integration of Official Statistics, European Conference on Quality in Official Statistics (Q2014), 2014.

[4] J. Thomas and K. Cook: Illuminating the Path: Research and Development Agenda for Visual Analytics, 2005.

[5] Wikipedia, the free encyclopedia: Hamming Distance:
(http://en.wikipedia.org/wiki/Hamming_distance).

[6] Wikipedia, the free encyclopedia: Edit Distance:
(http://en.wikipedia.org/wiki/Edit_distance).