



GNITED MINDS
Journals

*International Journal of
Information Technology
and Management*

*Vol. VIII, Issue No. XI,
February-2015, ISSN 2249-
4510*

DATA MANAGEMENT USING VISUALIZATION

AN
INTERNATIONALLY
INDEXED PEER
REVIEWED &
REFEREED JOURNAL

Data Management Using Visualization

Syed Ahad Murtaza Alvi

Lecturer, College of Computer and Information Sciences, King Saud University, Saudi Arabia

Abstract – *Researcher, Scientists is now faced with an incredible volume of data to analyze. To successfully analyze and validate various hypotheses, it is necessary to pose several queries, correlate disparate data, and create insightful visualizations of both the simulated processes and observed phenomena. Though numerous analysis and visualization tools have been built to improve the scale and efficiency at which analysts can work, there has been little research on how analysis takes place within the organizational context of companies, enterprises etc. This article is based on data analysis and visualization.*

Keywords: Data Science, Visualization, Data Flow

INTRODUCTION

Data science is the systematic study of digital data using scientific techniques of observation, theory development, systematic analysis, hypothesis testing, and rigorous validation. Data science entails an integrated view of the challenges of data and thus adopts a holistic approach to confronting these challenges.

The purpose of data science is to enhance the ability to collect, represent, measure, and apply data in order to describe, predict, and explain natural and social processes by:

- 1) Creating knowledge about the properties of large and dynamic data sets,
- 2) Developing methods to share, manage, and analyze digital data, and
- 3) Optimizing data processes for factors such as accuracy, latency, and cost.

The goals of data science include basic research and discovery, as well as applied research designed to inform decision-making for individuals, businesses, and governments.

REVIEW OF LITERATURE

- Existing visualization systems [1]

Visualization systems such as VTK [4] and SC Irun [5] allow the interactive creation and efficient manipulation of complex visualizations. These systems are based on the notion of data flows, where visualization is produced by assembling visualization pipe-lines out of

modules that are connected in a network. However, these systems lack basic data management capabilities and as a result, they have important limitations.

An important limitation of existing visualization tools is that they do not provide mechanisms for capturing provenance. Manually created captions and filenames are often the only provenance information available for an image. They also lack history management since a single instance of a dataflow is maintained; any changes to a dataflow are destructive. In particular, because there is no separation between the dataflow specification and its parameters, as the parameters are modified, the previous values are forgotten. This places the burden on the scientist to first construct the visualization and then to remember what values led to a particular image. As the dataflow evolves (i.e., operations are added, deleted or modified) no information is kept about previous versions. Another limitation of existing tools is that they do not provide support for comparative visualization. In particular, they lack the necessary infrastructure for properly supporting exploratory multi-view visualizations. The process required to create and compare a large number of visualizations is way too cumbersome. For example, executing the same dataflow with different parameters (e.g., different input data sets) requires users to manually specify all the parameters using a Graphical User Interface (GUI). Clearly, this mechanism is not scalable for generating more than a few visualizations. Finally, existing systems lack an optimization infrastructure. In particular, these systems may perform unnecessary and redundant computations while executing data flows.

VisTrails addresses these limitations providing infrastructure to support the interaction of the scientist with the visualization process. Our objective is to give scientists a dramatically improved and simplified process to analyze and visualize large ensembles of simulations and observed phenomena.

DATA VISUALIZATION

Data visualization is a general term that describes any effort to help people understand the significance of data by placing it in a visual context. Patterns, trends and correlations that might go undetected in text-based data can be exposed and recognized easier with data visualization software.

Today's data visualization tools go beyond the standard charts and graphs used in Excel spreadsheets, displaying data in more sophisticated ways such as info graphics, dials and gauges, geographic maps, spark lines, heat maps, and detailed bar, pie and fever charts. The images may include interactive capabilities, enabling users to manipulate them or drill into the data for querying and analysis. Indicators designed to alert users when data has been updated or predefined conditions occur can also be included.

Data exploration and visualization requires scientists to go through several steps. They need to select data sets and design complex data flows that apply series of operations to the data to create appropriate visual representations, before they can finally view and analyze the results. Often, insight comes from comparing the results of multiple visualizations. Unfortunately, today this process contains many error-prone and time-consuming tasks.

In addition, once a data product, e.g., an image, is generated, all the scientist is left with is the bitmap; if a detailed caption is not created, it may not even be possible to reproduce that image at alter time. As a result, the generation and maintenance of visualizations is a major bottleneck in the scientific process, hindering both the ability to mine and use scientific data.

The VisTrails system [2, 3] represents our initial attempt to stream-line the visualization process. Our long-term goal is to provide the necessary infrastructure to improve the scientific discovery process and reduce the time to insight. In VisTrails, we address the problem of visualization from a data management perspective: Vis-Trails manage the data and metadata of visualization products. By capturing the provenance of both the visualization processes and data they manipulate, VisTrails enables reproducibility and simplifies the complex problem of creating and maintaining visualizations. It also allows scientists to efficiently and effectively explore data through visualization: they can explore their visualizations by returning to previous versions of a dataflow (or visualization pipeline); apply a dataflow

instance to different data; explore the parameter space of the dataflow; query the visualization history; and comparatively visualize different results. Data management techniques used in many different components of the system are key to providing these functionalities, which have been absent in previous visualization systems.

VISUALIZATION TOOLS FOR DATA MANAGEMENT

Data visualization is a graphical representation of numerical data. The right data visualization tool can present a complex data set in a way that is simple to understand. Data tools enables users to create graphical and often interactive representations of data sets big and small can contribute greatly to improvements in corporate business intelligence (BI) efforts and organizational productivity. In the online survey of 210 BI practitioners and business users, 74% of the respondents said the influence of data visualization on attaining business insights within their organizations was “very high” or “high,” while another 23% said the technology had a “moderate” influence on the BI process.

The survey also indicated that the use of data visualization tools was helping to drive increased adoption of BI dashboards, typically the preferred medium for accessing and viewing charts, maps, graphs and other visuals.



Figure 1: Example of an automobile dashboard, displays key performance indicators (KPIs)

➤ Google Analytics dashboard [7]

The Google Analytics dashboard is a user interface to the company's Web analytics product. Google Analytics is a free service that provides statistics and basic analytical tools for marketing and search engine optimization (SEO) purposes. Within the dashboard, users can save profiles for multiple websites and either sees details for default categories or select custom metrics to display for each site. Available categories for tracking include content overview, keywords, referring

sites, visitor's overview, map overlay and traffic sources overview.

CONCLUSION:

In this paper we found that most business intelligence software vendors embed data visualization tools into their products, either developing the visualization technology themselves or sourcing it from companies that specialize in visualization.

REFERENCES:

1. Steven P. Callahan Juliana Freire Emanuele Santos Carlos E. Scheidegger Claudio´ T. Silva Huy T. Vo "VisTrails: Visualization meets Data Management"
2. L. Bavoil, S. Callahan, P. Crossno, J. Freire, C. SC Heidegger, C. Silva, and H. Vo. Vistrails: Enabling interactive multiple-view visualizations. In IEEE Visualization 2005, pages 135–142, 2005.
3. S. Callahan, J. Freire, E. Santos, C. Scheidegger, C. Silva, and H. Vo. Managing the evolution of dataflows with vistrails. In IEEE Workshop on Workflow and Data Flow for Scientific Applications (SciFlow 2006), 2006.
4. Kitware. The Visualization Toolkit (VTK) and Paraview. <http://www.kitware.com>.
5. S. G. Parker and C. R. Johnson. SCIRun: a scientific programming environment for computational steering. In Supercomputing, 1995.
6. <http://whatis.techtarget.com/definition/data-visualization-charts-graphs-dashboards-fever-charts-heat-maps-etc>
7. <http://searchbusinessanalytics.techtarget.com/definition/Google-Analytics-dashboard>