SuperViewer: An Interactive Visual Interface to Explore the Top500 List

Leonidas Deligiannidis  
Wentworth Institute of Technology  
Computer Science  
550 Huntington Ave.  
Boston, MA  
deligiannidis@wit.edu

Erik Noyes  
Babson College  
Entrepreneurship  
231 Forest St.  
Babson Park, MA  
enoyes@babson.edu

Hamid R. Arabnia  
University of Georgia  
Computer Science  
415 GSRC, UGA  
Athens, GA  
hra@cs.uga.edu

ABSTRACT
In this paper, we describe SuperViewer, an interactive visual interface which facilitates exploration of the “Top500” list, the biannual list of the world’s most powerful supercomputers (1993-present). SuperViewer is both intended as a visual knowledge discovery tool for computer science students as well as a dynamic presentation tool for supercomputing experts to represent and debate points of view on critical technological and entrepreneurial developments in the industry, including the industry’s past and projections for future innovation. The tool offers four ways to view and interact with historical data to explore and represent the rapid pace of innovation and competition in the industry.

Keywords: supercomputing, information visualization, education, visual knowledge discovery, visual interfaces

1. INTRODUCTION
By all accounts, the historical pace of change in the supercomputing industry has been staggering. In this paper, we describe an effort to build an interactive visual history of entrepreneurial activity in the supercomputing industry, considering computers’ speeds, manufacturers, and design architectures to represent and explore broader technological disruptions shaping the industry. We present ongoing work on an interactive visual interface we name SuperViewer, which offers four ways to view data from the Top500 list for all the periods of the biannual list’s existence (1993-2012).

SuperViewer has two intended purposes. First, it offers those teaching about supercomputing new tools to visually present and characterize historical developments (e.g., to new students of computer science). Second, it allows field experts to visually represent and debate points of view on critical entrepreneurial and technological developments in the industry, including the industry’s past and projections for future innovation.

As an interdisciplinary project, our aim is to exploit domain knowledge in information visualization [1][2], interface design [3][4], and visual knowledge discovery [5] to provide windows into the concept of creative destruction [6][7][8]. Creative destruction is the foundational idea in entrepreneurship that technological disruption, market upheaval and specifically new venture successes and failures drive industry
evolution and economic development [6]. The development of SuperViewer is part of broader effort to develop new, interactive teaching tools to highlight the dynamism of industries and entrepreneurship phenomena.

2. THE SuperViewer
We designed this application entirely in java. This enabled us to place the application on the web for researchers, educators and students to easily access it using Webstart technology. The link to the home page of the project where the application can be launch is: http://faculty.cs.wit.edu/~ldeligia/PROJECTS/Top500/index.html. The data was retrieved from top500.org as individual files; we extracted all critical fields and compiled one single data repository.

Top500.org publishes its data twice a year, once in June and once in November. The data is available from June 1993 to November 2012. The data are not consistent across all years and the format is not the same in all data files, so there was a considerable amount of data cleaning/data checking. No unique identifier is available for supercomputers listed on the site, thus we created a unique identifier as the concatenation of the “ComputerName” and the “SiteName” fields, which effectively served this function. For example, we found that the same Top500 computer name in one year could be miss-spelled, parts of it capitalized, containing or missing spaces or underscores, etc. The data is likely 97% clean at this point but additional checks are needed.

Description of what fields are provided by top500.org is given on their web site. Two of the fields that we use in all four visualization tools are the $R_{\text{max}}$, which is the maximal LINPACK performance achieved by a supercomputer, and $R_{\text{peak}}$, which is the theoretical peak performance. Both fields are measured in GFlops. Our application consists of four visualization methods and they are described in the sections below.

2.1. Visualizer “Ranking”
The Visualizer “Ranking”, shown in figure 1, visualizes the top 500 supercomputers per year. For each year, we plot the data of all 500 supercomputers based on their $R_{\text{peak}}$ and $R_{\text{max}}$ values, which are measured in GFlops. There are two ways of plotting the data, the first is (max-per-year) and the second is (max-overall). In max-per-year mode, the maximum scale on the Y axis is the top $R_{\text{peak}}/R_{\text{max}}$ value of the number 1 or fastest supercomputer for the selected year.

In max-overall mode, the maximum value on the Y scale is the maximum value of the top supercomputer of all time/all years. This way, one can compare the top 500 computers within a year (max-per-year), or one can compare the overall speed (max-overall) of the computers by keeping the scale the same for uniform comparison across all years. We also provide the ability for a user to display the Moore's law curve for the past data and make predictions for future data. The user interacts with this visualizer using a Combo box to select the type of visualization, and a slider. The Combo box, shown in figure 2, is used to select the $R_{\text{peak}}/R_{\text{max}}$ variables for visualization as well as the mode (max-per-year, and max-overall) of visualization. The slider is used to select the year of the dataset to be visualized.
Figure 1. The Visualizer “Ranking”, visualizes the top 500 supercomputers per year based on their Rpeak and Rmax values.

Figure 2. Variable and mode of visualization in visualizer “Ranking”.

2.2. Rank Animator
The Rank Animator, shown in Figure 3, is a highly interactive visualization tool built on top of a spring-embedded algorithm that animates the Top500 ranking data as it changes over time. The user can select the top N supercomputers to visualize. Then the user can use a slider to retrieve and visualize these top N supercomputers for specific years. The left and right arrow-buttons can be used to move forward or backward one year at a time. Or the user drags the slider instead to adjust the viewed time frame.

This spring embedded visualizer automatically arranges the top N supercomputers from left to right based on their rank (Rmax value). The user can double click nodes to drill down and get more
information about a particular supercomputer of interest. As the user clicks the buttons or drags the slider, the new ranking of the supercomputers is shown. The spring-embedded algorithm animates and rearranges the graph for the user to observe how the ranking is modified between different times (top 1-20 or as otherwise specified). Each node is labeled with its name and on top of it, within a circle, showing its ranking in the selected time period. The visualizer also annotates in color new supercomputers that were not part in the previous visualization frame as yellow. Bluish nodes indicate that a particular computer was at the same rank in the previous frame as the current frame. Green nodes illustrate that these nodes were part of the graph in the previous frame, but that their rank has changed in the current frame. The visualization in Figure 3 below shows green nodes (e.g., nodes 3-5), as well as notations for changes in rankings for individuals computers. For example, the “4->5” annotation below node 5 indicates that this computer used to rank number 4 in the previous frame, but now it ranks number 5.

![Image of visualization tool](image)

**Figure 3.** The “Rank Animator” is a highly interactive visualization tool that animates and arranges, in rank order, a sub-list of the Top500 supercomputers as they change over time.

2.3. **Visualizer “Track #1s for all years”**

This visualizer is used to visualize the history of all number-one spot supercomputers (e.g. the Top 1 in the Top500 list). It visualizes, in a single window, all number one computers and their ranking as the years progress. A number one computer in 1995 becomes number 4 the next year and number 20 the year after and so on. The green circles indicate a new super computer introduced at the specific year. Its history is a line moving downwards as the years pass by, as shown in figure 4. The zoom-factor slider is used to zoom in the data to see the history in more detail as shown in figure 5. A number above a green node indicates the number of cores/processors for this particular supercomputer.
Figure 4. The visualizer “Track #1s for all years” visualizes the history of all number-one spot supercomputers; all number one computers and their ranking as the years progress.

Figure 5. The zoom-factor slider is used to zoom in the data to see the history in more detail.

2.4. Visualizer “Track N starting in a year”
This visualizer is similar to the previous visualizer in regards to visualizing the downfall of top supercomputers of a specific year. In this visualizer, a user selects the top N supercomputers of a selected year. The visualizer then displays their ranking as the years progress, as shown in figure 6. The ranking of the supercomputers is based on the Rmax value. Moving forward in time, we see how fast highly-ranked supercomputers become less powerful as new faster supercomputers replace them. An alternative to this type of visualization could have been a matrix-based representation, which is suitable for large and dense graphs [9]. However, because of the simplicity of the data we found that matrix-based representation does not add any value to this tool.
3. TRACKING Roadrunner

Roadrunner earned the number one slot in the top 500 list of supercomputers in June and November of 2008 and once again in June of 2009. In March of 2013, Roadrunner was declared obsolete and was scheduled to be dismantled [10][11].

This IBM-built supercomputer was designed to model the decay of the US nuclear weapons arsenal. It is still powerful enough to hold the 22nd slot in the list of the top 500 supercomputers. However, it is considered an energy inefficient machine and that is the main reason for considering it obsolete.

Figure 7 is a compact snapshot of the Visualizer “Track #1s for all years”. It displays the number of processors, and the decline in ranking since June 2008, from the number 1 rank down to number 22. Figure 8 consists of 10 snapshots of the Rank Animator. This figure visualizes Roadrunner from June 2008 to November 2012. We annotated (with current rank in red circles) the figure to easily track Roadrunner through history and its rank changes.
Figure 8. Ten snapshots of the Rank Animator while visualizing the history of the supercomputer Roadrunner.

4. FUTURE WORK
Future work aims to expand the range of data and views examinable in SuperViewer. We are in the process of adding other data fields and visualization possibilities to the tool, including the ability to superimpose information about computer architectures, manufacturers, international site locations, public or private ownership and other technical specifications for 1993-2012. Data has been collected but we need
to work on techniques to visualize and display this information. Future opportunities also include adding statistics to presentations, for instance capturing average computer performance by year (e.g., Rpeak or RMax) as well as other informative descriptive statistics. Overall, the aim is to create, test and offer a user-friendly viewer that enables students and field experts alike a tool to explore and represent data to characterize both evolutionary and revolutionary changes in the industry.

5. CONCLUSION
SuperViewer offers supercomputing experts and students an interactive visual tool to present historical developments in the industry. Through simple controls, the viewer provides access to the dynamism of the industry, considering both entrepreneurial and technological developments—all with a streamlined interface that facilitates data querying and representation. While the development of SuperViewer is in progress, a key aim of the project is ultimately to engage global supercomputing experts in order to identify other design and interactivity dimensions for the tool which will support the representation and exploration of key changes in the industry.

REFERENCES