Querying Web Site Visitor Trend Data with Coordinated Nested Bar and Pie Charts

Mark Sifer
School of Computer Systems Engineering
University of New South Wales
Sydney, New South Wales, Australia
msifer@cse.unsw.edu.au

Abstract

Bar charts and pie charts are traditional ways of displaying trends and proportions in a data set. However, static charts don't scale well. We introduce a technique which does scale, an interactive browser design which uses aggregation and zooming to support bar and pie charts with thousands of items. Further, once a chart has many elements a user needs to be able to explore subsets of the data. We use a second view of the data, shown as a second chart as a scalable filter for selecting subsets. We present our design and its implementation in the SGViewer tool, and demonstrate with visit data taken from a web site log.

Keywords: User interface, bar chart, pie chart, hierarchy visualisation, multiple views, query.

1 Introduction

Bar charts are often used to show trend data, with relative bar heights showing how a value changes over time. When chart values are drawn from a large dataset users may wish to explore trends for subsets of the data in addition to seeing the overall trend. Subsets need to be identified in some way. Often a dataset will have multiple dimensions. For example in a dataset of visits to a web site, each visit has a date, a number of pages visited count, a refer domain, a visitor domain, a list of downloads that the visitor made during their visit and so on.

How to subset data depends on the kind of data. If it is unordered, a subset is selected one-by-one. If it is ordered, a subset can be selected with values ranges using perhaps a slider or a text form to enter start and end interval values. If the dataset is partially ordered and forms a tree, a subset can be selected with a tree selector through a series of subtree selections. When sliders are used for ordered data, both linear and non-linear scales may be used depending on the kind of intervals users are likely to need. When a tree selector is used some kind of scaling to assist user selections may also be needed.

We present our solution for a pie chart tree selector integrated with a bar chart next.

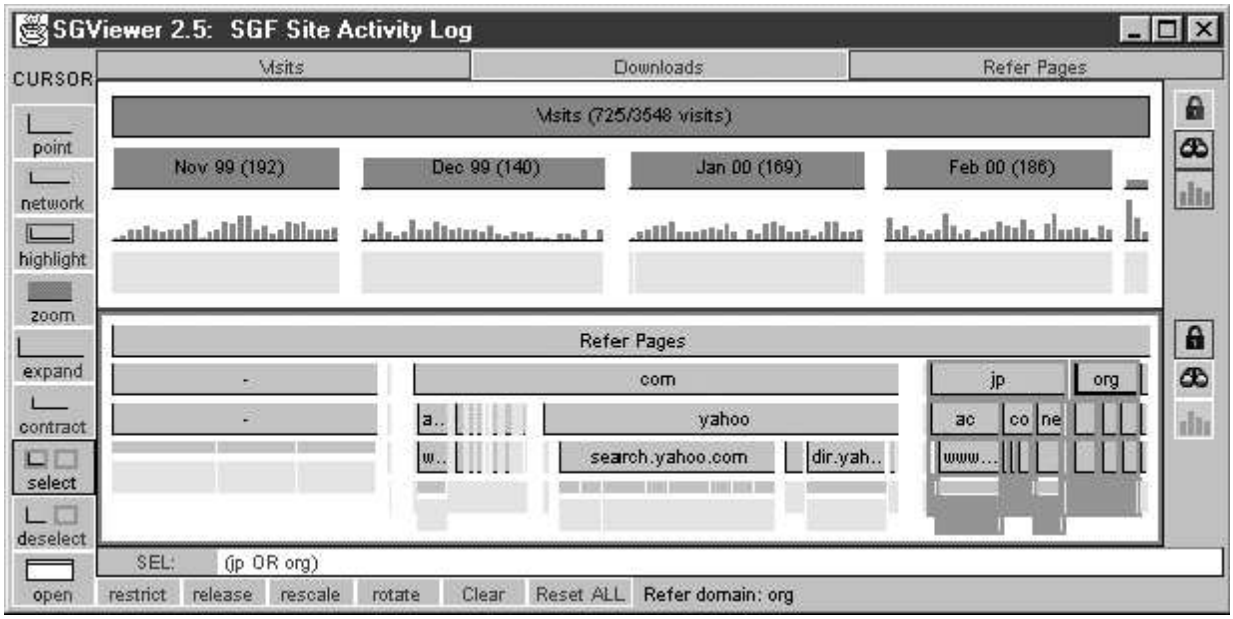

Figure 1: SGViewer browsing a web log

2 Exploring The Data Set

Figure one shows SGViewer being used to browse a web site log data set. The data has been organised into two trees: a calendar and refer page tree. These trees overlap, sharing the same leaf nodes; the individual visits to the

SGF web site from unique IP addresses. The lowest layer of each tree contains the same 3548 visits. Individual boxes for each visit can not be seen as they are too thin, rather they are greeked together into lightly shaded contiguous areas for improved performance.

The calendar tree is a nested bar chart. Numeric data annotates the higher level boxes, so there were 186 visits in February that came from the 'jp' or 'org' high level domains. The fill of each box is varied to show the relative number of leaves (visits) which have been selected in the refer domain panel. The layout has been constrained so that each individual day node on the third level of the calendar tree has equal widths, ensuring this level gives a daily visitor trend for the selected visits. The second level summarises the third level, showing the monthly visitor trend. We can see that after November 1999 the number of visitors declined but then slowly recovered, and in the last three days of March there was a surge of visitors.

The refer tree is a nested pie chart in the horizontal dimension. It shows where visitors to the SGF site have come from. The relative width of each box corresponds to the relative proportions. The most common referrer is yahoo.com accounting for about 30% of refers, with most of these coming from keyword searches and some from yahoo's directory listing. After yahoo the next most common refer is no referrer, signified by the dash. The refer tree is a tree that has been constrained so that each leaf node box has equal width and inter-box spacing is kept to a constant. The width of non-leaf nodes is then proportion to the number of leave nodes it has, in our example this means the number of visit nodes at the bottom.

A user browses different trends for site visits by making selections in the refer domain tree. Selected subtree's are shown with a blue border. On startup no selection has been made and the calendar shows the trend for all visits. As the user makes refer selections by clicking and shift clicking refer domain subtrees (selecting one or more subtrees) the calendar changes to show the trend for the selected visits. The pie chart proportions of the refer tree which identify the main contributors to the overall trend can help guide which refer domain should be explored first.

To see a subtree in more detail zooming can be used. To refine a selection, restriction can be used. Figure two shows the result of restricting the refer panel. The refer tree now only contains 'jp' and 'org' subtrees. If the 'xml.org' subtree was selected next the calendar would change to show the trend for visits from 'xml.org' only. Both zooming and restricting are done with animated transitions. The coupling between the two trees is symmetrical, so the role of the calendar and refer domain could be reversed, with the user making selections in the calendar and the refer domain tree automatically changing (with an animated transition) to show the proportions for the selected dates. For example November could be selected first, then while using a cursor key to change months, the refer domain tree proportions will smoothly vary as the selected month changes.

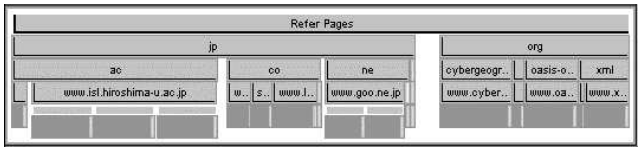

Figure 2: restricting the refer tree to jp and org

3 Related Work

Treemaps (Shneiderman 1992) is a 2D nested visualisation, but because it nests in 2D it is less suited to acting as a filter; for restricting to selected subtrees would not preserve geometric proportions. Dynamic queries (Ahlberg and Scheiderman 1994) use multiple sliders to restrict the resulting data shown in the primary visualisation panel, such as a scatterplot. (Dix and Ellis 1998) presents a pie chart in which segments can be exploded under the control of an text tree outliner view, allowing access to multiple levels of the data.

4 Conclusion

We have presented a design and its implementation for the exploration of trend data using a nested bar chart for displaying results and a nested pie chart for making selections. The calendar nested bar chart aids reading of trend data by providing detailed trends together with summarised trends. The refer nested pie chart provides a visual breakdown of a data set (visits to a web site) and provides a selector suited to subseting data which has been organised into a classification (visits organised by refer domain). Our approach can also be applied to other datasets, which benefit from aggregation and multiple classifications, as SGViewer operates on arbitrary Structured Graph Format (SGF) XML files (Liechti, Sifer and Ichikawa 1998).

5 Acknowledgments

We wish to thank Hiroshima University where the viewer was initially developed, for granting the author use of the viewer and its source code allowing its refinement. Thanks to Michal Jacovi and Sarai Sheinwald for an improved look and feel, and to Einat Amitay for a helpful review.

6 References

AHLBERG, C. and SHNEIDERMAN, B. (1994) Visual information seeking: tight coupling of dynamic query filters with starfield displays, *Proceeding of ACM CHI*, 313-317, ACM Press.

DIX, A., and ELLIS G. (1998) Starting Simple - adding value to static visualisation through simple interaction, *Proceeding of Advanced Visual Interfaces*, 124-134, ACM Press.

LIECHTI O., SIFER M. and ICHIKAWA T. (1998) Structured Graph Format: XML metadata for describing web site structure, *Proceeding of the 7'th World Wide Web Conference*, 11-32, Elsevier.

SHNEIDERMAN, B. (1992) Tree visualization with treemaps: a 2-d space-filling approach, *ACM Transactions on Graphics*, 11(1), 92-99, ACM Press.