ADAPTING THE TILEBAR INTERFACE FOR VISUALIZING RESOURCE USAGE

Larry Reeve
CIMS Lab, Inc.

Visualizing resource use is important for discovering patterns of system component usage over time. This paper describes how the TileBar interface, originally developed for use in information retrieval, can be adapted for visually analyzing resource usage. The 1-dimensional TileBar can show up to four attributes simultaneously, allowing analysts to quickly determine what is being used, how much, when, and by whom. TileBar displays can also be used as a navigation interface for drilling down into resource usage detail.

Introduction

Information visualization, through the graphical presentation of data, is useful for compactly displaying large amounts of information within a small display space. Patterns in data are often revealed when they are shown in a graphical display. These patterns are not obvious, or even evident, when displayed in other forms, such as plain text displays. The human visual system is a high-bandwidth facility [CAR99], and so it makes sense to take advantage of this facility when handling large amounts of data. This is particularly true in the area of resource management, where the data collected on a daily basis is often great, spans many resource types, and is difficult to make sense of without techniques such as data aggregation that reduce the amount of data handled and analyzed.

Aggregate information can also overwhelm users, particularly when presented in text-oriented reports, where each resource usage is displayed line-by-line. Such reports are useful for seeing aggregated usage information on an account-by-account basis, as shown in Figure 1.

		Total All Months	9/2004
AAAAAALL - Larry			
	640 - MS Exchange Bytes Sent	27,399,318	27,399,318
	641 - MS Exchange Emails Sent	83	83
	642 - MS Exchange Bytes Received	12,275,671	12,275,671
	643 - MS Exchange Emails Received	202	202
AAAAAARB - Robert			
	640 - MS Exchange Bytes Sent	7,007,730	7,007,730
	641 - MS Exchange Emails Sent	199	199

Figure 1: Sample text usage report by account.

In this type of report, it is difficult to see how usage is changing over time within an account, unless a report is run for each period of interest, as shown in Figure 2. This type of report is again narrowly useful, but is difficult for an analyst looking for usage trends to easily discern the patterns. For example, rapidly providing answers to the following types of questions is difficult: What are the peak periods for email usage? Is usage trending up or down or staying the same? Are the sizes of email messages sent increasing, decreasing, or staying the same? There are many similar types of questions that are difficult to answer using only text reports. Trends and patterns are hard to identify in textual form.

		1/2004	2/2004	3/2004
BBDDDDLL - Larry	782 - Unix Oracle Messages Received	197	161	120
	833 - Unix (512-byte) Block weeks	29,760,370.18	17,356,064.09	28,914,236.41
	834 - Unix Filesystem Size (512-byte	403,549,340.00	251,879,808.00	384,448,128.00
	835 - Unix Filesystem Blocks Used	202,832,645.00	128,610,657.00	195,591,765.00
	836 - Unix Filesystem Number of Files	2,223,234	1,392,672	2,127,047
	837 - Unix Filesystem Size (GB days)	0	0	133
	838 - Unix Filesystem Used (GB days)	0	0	67

Figure 2: Sample text usage report by account over 3 periods.

Commonly used business graphical displays, such as bar charts, can show the text-oriented report information in a visual format. The information content, however, is the same. The graphical presentation does reduce the cognitive load on the user, since visual aids, such as the height of bars in a bar chart, offload work to the human perceptual system [CAR99]. Typical business visualizations are

also usually limited to three dimensions at the maximum in common software tools such as Microsoft Excel. If a resource usage analyst wants to look at user account, time periods, usage amounts, and usage-type information simultaneously, three-dimensional tools become inadequate.

A challenge in visualizing resource usage for analysis is finding ways to compactly display aggregated data with more than three attributes while also reducing (or at least not increasing) the cognitive load on the analyst. A larger number of attributes allows for more immediate information so that 1) there is no shift of focus away from the visualization to some other information display to get required information (disruption elimination), and 2) pattern discovery is potentially enhanced.

One solution to this problem may be found in bar displays. In this paper, TileBars are examined and adapted, but they are only one type of bar display. Another type of bar display is the ValueBar, which uses a single, completely-visible bar for each attribute, and displays resource value indicators within the bar using size as an indicator of the resource value [CHI92,CHI91].

TileBar Interface

The TileBar is a visual interface developed in the information retrieval field to compactly show quantitative attributes of a document. The TileBar interface is used as a visualization paradigm to aid users in determining the potential relevancy of a document within a set of documents [HEA95]. A user issues a document retrieval search using various query terms, and the result of the search is a listing of documents along with a TileBar display next to each document. The display shows the frequency of each query term within each section of a document, the length of a document, and the document name. A user looking at a TileBar can immediately see which documents have used the requested query terms the most, as well as the position in the document of the frequency counts. This allows users to see not only which documents may be most relevant, but also at which positions within each document may be most relevant [HEA95]. Navigation is eased by allowing the user to immediately navigate to the section of the document where the terms occur.

Each bar display in a TileBar display is composed of one or more individual tiles. Each tile represents some section of a document, such as subtopic, page, and so forth. Tiles are coded using color intensity (color or grayscale can be used). Darker intensities indicate larger term frequency counts. The tiles for a particular term are combined in a non-overlapped manner so that they form a horizontal bar display. Several bars can be stacked horizontally to represent a document. The result is a display that reflects 1) relative length of a document, 2) term frequency, 3) distribution of terms within a document, and 4) exploits the pattern-recognition capabilities of the human perceptual system; that is, patterns can be quickly scanned and deciphered [HEA95].

Figure 3 shows an example TileBar display for a single document using three search terms. There are three bars, one for each term. Each tile represents a page. The term frequency counts are coded using grayscale intensities, where the higher the intensity, the larger the frequency count of the query term. In this example display, it is easy to see that all three query terms have a higher frequency count towards the middle of the document. Users can navigate to the middle section of the document in order to make a determination of the document's relevancy. The common alternative interface is for the user to display the full document and then search within the document for the search terms. The TileBar, then, serves as a sort of map to guide the user to potentially relevant areas within a document.

Figure 3: TileBar display for single document with three search terms (HEA97).

There are several primary components of the TileBar that can be identified. Each bar is associated with a single term. The length of a bar for a term is determined by the length of the document using some metric, such as page count. The term frequency is categorized rather than absolutely displayed. The categorization is represented using color intensity (either color or grayscale). Finally, all search term information for a particular document can be shown at one time by stacking the bars horizontally, one on the other. TileBars are considered one-dimensional displays that display information along the X-axis [CAR99]. The power of TileBars is their simplicity of design enables the display of multiple attributes of information within a single dimension. This ability allows for a large degree of information to be displayed compactly, and without resorting to more complex 2D and 3D visualization methods.

Hearst and others [HEA96] solicited qualitative feedback on the use of TileBars in information retrieval in order to gain some sense at the effectiveness of the TileBar display interface. Their findings were that all users in the study found the TileBar useful to some degree. Some liked the TileBar because it provided a global view of a document while also allowing for drilldown into particular sections of a

document. This is no surprise to those who subscribe to the Visual Information-Seeking Mantra proposed by Ben Shneiderman: “overview first, zoom and filter, then details-on-demand” [SCH96]. Users also said they used the TileBar to quickly discover information about a document; in other words they used the TileBar pattern to make a decision about whether or not the document was likely to be relevant without reading any of the document text. Other users wanted the TileBar sizes to be larger. Presumably this is to allow the individual tiles to be expanded in width so that sections of a document are more distinguishable. Wider tiles in turn increase the length of the bar. Users also found the TileBars more useful when a query result set has been narrowed down to a manageable level. For example, if a document result set is very large because the query did not provide enough focus on the requested topic, the TileBars were not as useful. If the document result set was small, the TileBars help the user decide which entries in the returned set may be most useful. Most users said they used the TileBar at the end of their search process, suggesting that while the TileBar reduces cognitive load in a small result set, in a large result set the TileBar was not of very much help in narrowing down potentially relevant documents. No known empirical user studies of the TileBar interface have been conducted.

TileBar Adaptation

While the TileBar was originally developed for use in another field, information retrieval, the TileBar is constructed from various components that can be thought of as generic: tiles combined into bars, and bars stacked together to represent grouping. Using these generic parts, the question is whether or not resource usage information can be displayed using similar techniques and with similar results. An adapted TileBar visualization would be useful to analysts in discovering patterns of resource usage over some defined period.

In the first attempt to adopt the TileBar for resource usage display, called the Resource Usage TileBar, the construction used the same format as the original TileBar: tiles of even size combined into bars of varying lengths, and the bars then stacked horizontally to show relatedness. The individual tiles showed a usage category for each period using a color intensity scheme, where darker intensities represent increased usage. The category represented was either “no usage” or “low”, “medium”, or “high usage”. The initial coloring for each category was done using grayscale. The length of individual tiles was also of a reasonable width so that each tile can show its usage category without the user straining to see it. The width size was not precisely defined, but was larger than a few pixels. An adequate width size is easy to recognize when the tile is large enough to distinguish when combined with other tiles, but at the same time the width size is imprecise to specify or recommend, because it depends on the overall size of the visualization rendered. The grouping of usage tiles into bars represented a particular resource, shown using a rate code. The group bar lengths were determined by averaging usage categories. The bars for each resource user were stacked (grouped together) based on the resource user that they represent. A primary change from the original TileBar design that was tried was to set the length of each bar to a variable-length, rather than having each group of bars of variable-length. The intent was to show the relative usage of each resource across multiple resource users. However, this changed the width of each tile, resulting in tile bars that looked disjoint, rather than part of a group, and so was rejected. Labels were also added to show the resource user, using an account code, as well as for resource type, using a rate code.

An example of the initial application of the TileBar to resource usage measurement is shown in Figure 4. The user accounts, shortened for presentation readability, are indicated as AA, BB, and CC. The resource types are indicated using the rate codes EXEMST – Email Messages Sent, EXEBYST – Email Bytes Sent, EXEMRCV – Email Messages Received, and EXBYRCV – Email Bytes Received. Each resource type is shown using a single bar, and all the bars are combined to form one large bar for the user account. Each tile in a bar represents one time period (in this case, one month, so each bar is made up of twelve individual tiles to show one year). The number of time periods can be adjusted to show more or fewer periods. The intensity shading of each tile gives the relative resource usage. The length of each user account bar gives the relative usage of all resources as compared to other users.

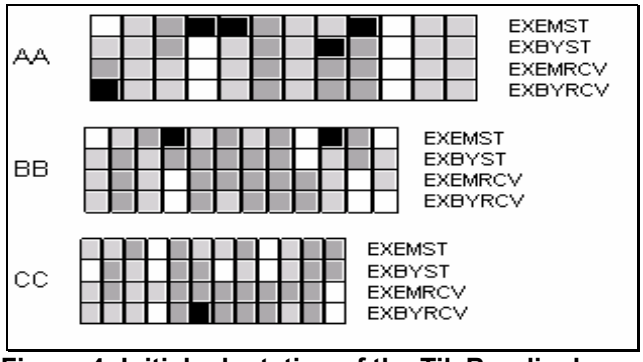

Figure 4: Initial adaptation of the TileBar display for resource usage showing resource user account, twelve periods, usage class, and resource type.

After a sample Resource Usage TileBar was constructed, feedback was solicited from several individuals familiar with resource accounting reporting

and who rely on such reporting on a daily basis. Initial results showed the generated display was not considered complete. For example, the varying lengths of the bars, while they showed relative usage information, made the comparison between time periods difficult to analyze. The eyes are continuously searching for the same periods between multiple resources in order to provide comparisons. The visual search becomes more of a challenge when trying to compare the same resource across two different resource users, since the distance between the two bar displays is greater. Labels for each time period were requested. A final suggestion was to use a colorized display for each tile rather than grayscale. The implication is that visualizations based on grayscale tend to look dated when placed next to other common business visualizations, such color bar and pie charts. The Resource Usage TileBar was modified to include these changes. The result of the TileBar adaptation is shown in Figure 5.

There are several changes from the original TileBar design: labels, tile widths, and bar size. The most significant change is making the tile bars even-sized across all resource users, which forces the tile bars in the entire display to be of the same length. This change results in the relative usage information of a resource being lost from the display, which is a negative for the adaptation. However, further feedback from the original group suggested that patterns of resource usage between and across time periods are easier to discover with evenly-sized tiles. The tradeoff of losing overall relative resource information as compared to gaining potentially increased resource pattern detection was deemed worthwhile. Also, in implementations, users can be presented with an option to show one or the other, depending on the analysis need.

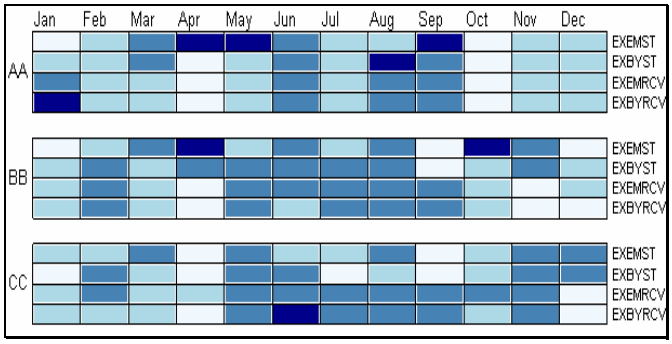

Figure 5: Adapted Resource Usage TileBar display showing 4 attributes simultaneously: resource user account, twelve periods, usage class, and resource type.

The addition of labels helps clarify the multiple dimensions of information that users are viewing. The TileBar is a compact design, and through feedback it was discovered that labels serve as a convenient reference. Labels were included in the original TileBar design only for the document identifier, and not for the query terms or document section. The difference may be that the TileBar was intended as a navigation aid to document contents, while the Resource Usage TileBar is designed for pattern analysis. Resource Usage TileBar users may spend a longer time interpreting the patterns displayed in order to find meaning, as opposed to quickly scanning a TileBar to navigate to particular sections of a document.

Results

The Resource Usage TileBar adaptation is demonstrated through a use case. Assume there is a resource usage analyst interested in looking at email usage from the prior year. There are three resource users of interest to the analyst, coded as AA, BB, and CC. The goal of the analyst is to find answers to several questions: What are the peak periods for each resource user? Is email usage trending up or down for total emails, emails sent, and emails received? The analyst has access to an organization-wide database containing resource usage information. The database can be consulted about detailed information on the number of emails sent and received, as well as the size of messages sent and received. The resource database is used to summarize and categorize the detailed usage data, and then use the results to generate a Resource Usage TileBar visualization, as shown in Figure 5.

Figure 5 shows all of the information requested by the analyst. Information for all three resource users is shown simultaneously. Each resource user has four tile bars stack together. Each tile bar represents one email resource over a time period of 12 months. The tile bars are labeled with their internal code: EXEMST is number of emails sent; EXEBYST is the number of bytes sent; EXEMRCV is the number of emails received; and EXBYRCV is the number of bytes received. Each tile is color-coded using various intensities of blue, where a darker intensity indicates a larger amount of usage. White indicates no usage.

The following observations can be made from Figure 5 for each user. User AA tends to send more emails than received, although the size of the messages is small as compared to other users. The peak months are in April, May, and September. The heaviest usage is in the period June to September. Email usage for send and receive peaks in September, then declines to a moderate level by the end of the year.

User BB sees the heaviest usage for send and received in the time period of May through August, although February is also a peak period. Email usage for send and receive peaks in the August-September

timeframe, and tends to decline through the rest of the year.

User CC has fairly steady usage in the moderate category from May until the end of the year. The primary use of email for user CC seems to be in receiving emails rather than sending. The size of messages received also seems to be fairly high, as compared to other email resource users. The trend seems to be increased email usage as compared to the start of the year.

The Resource Usage TileBar display can also be used as a high-level navigation tool for drilling down into resource usage detail. A user interface can be constructed so that each tile is selectable. If a user selects a tile, a request to return all of the detail records for a particular user/resource/period combination is generated, thus providing a very simple query interface into the resource usage database. For example, in Figure 5, selecting the tile corresponding to resource user "AA", time period "Jan", and rate code (resource type) "EXEMST" will return all of the individual detail records that make up the tile. This fulfills the Visual Information-Seeking Mantra of "Overview first, zoom and filter, and then details on demand" [SCH96]. That is, users gather an overview of the data, select their area of interest, and can then drilldown to find details.

To contrast charting with the Resource Usage TileBar, Figure 6 shows a column chart displaying the same information as Figure 5, but for a single resource (emails sent) and a single user (AA). Note the resource type is defined in the chart title, and not in the main graphic. It is possible to add additional users, represented by additional bars with different colors. It is not possible to add additional resources.

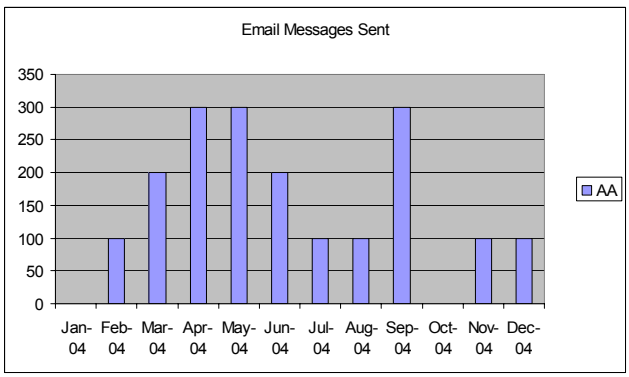

Figure 6: Column chart of resource usage for a single user over twelve periods.

To show multiple resources, the graph can be rotated so the title shows a single account code and the bars show resources rather than users. In this case, only a single user with multiple resources is shown, as opposed to a single resource with multiple users. As seen in Figure 5, the Resource Usage TileBar requires no adjustments to show both multiple users and multiple resources simultaneously.

Implementation

It is expected that implementations of the Resource Usage TileBar will be either as a user library, or incorporated into vendor products as a visualization feature. For example, vendors can offer the Resource Usage TileBar as an additional selection along with bar and pie chart selections. Users wanting to implement their own version of TileBars can look to a pre-built library implementation, or even roll their own. Since a TileBar is composed of tiles linked into bars, and then bars stacked together, the implementation is relatively straightforward. The prototype developed used less than 100 lines of Microsoft .Net C# code on the Windows platform to perform the line drawing and tile shading.

As with any visualization, the resulting quality is largely determined by the underlying data. Users need to decide what types of information are important to analyze. The TileBar is fed data that will allow it to be used in several scenarios. Examples include single account, single resource; single account, multiple resources; multiple accounts, single resource; and multiple accounts, multiple resources. The single account scenarios allow for comparison within one or more resources (trending, period comparisons, and so forth). The multiple account scenarios allow for comparisons within resources as well as between accounts across multiple periods. A key consideration for effective comparisons is deciding how to shade the individual tiles. Tiles show the usage category. This category may be determined within a single account across all periods. It could also be relative usage across all accounts and periods. Implementations must make it clear to the user what the usage category is measuring, as the TileBar is flexibile enough to allow for displaying many different usage scenarios.

Conclusion

In this paper, the TileBar visualization, first developed for presenting term frequency counts for retrieved document sets, is adapted to the field of computer measurement to show resource usage information. The Resource Usage TileBar is the name of the adapted TileBar.

Several changes are required to adapt the original TileBar display in order to show resource usage information. First, the size of each tile needs to be large enough for users to be able to easily distinguish the time periods within each bar. Resource Usage TileBars can be expected to have larger tiles than the

bar displays based on the original TileBar interface. Second, labels are added as a visual reference for identifying the four attributes displayed on each one-dimensional bar. Labels are not included on the original TileBar, but user feedback indicated that they improved readability by providing visual clues as to the identity of each of the attributes, so that users do not have to remember where each attribute is. Interestingly, there was no request to provide a legend for the color coding of each tile. The use of color intensity seems to have an intuitive appeal. The final change is the length of all bars are kept to the same width, both in the display within a bar group (resource type) and across bar groups (resource users). Constant bar widths enable easier comparison of time periods, although at the loss of indicating relative usage of a user of a particular resource.

Future work will include user studies of the Resource Usage TileBar interface. It is important that empirical evidence of user experiences be collected and analyzed. Such evidence can reinforce or be the cause for further analysis of the qualitative feedback already collected. The original TileBar was subject to qualitative analysis as part of a larger study, and was largely accepted by users. The Resource Usage TileBar interface has also had some qualitative feedback, but has not been subjected to a comprehensive study. It is expected that the Resource Usage TileBar will be further refined based on additional user feedback. Additional work will also be done to link individual tiles to their associated detail data, enabling users to perform drilldown interactively.

Finally, bar displays, such as ValueBars and TileBars, show that there are alternatives to the traditional ways for displaying information for analysis. As Card et al suggest, it is possible for information visualization to enable users to think by using the human visual system [CAR99]. Information visualization displays often reveal patterns that are not always evident by just looking at data [CAR99]. It is often useful to move from the world of text reports and business charting to the world of information visualization, where analysis using visual methods can often lead to substantial insights. As information systems become more complex and better instrumented, the means for analyzing the vast amount of data produced must also be extended.

References

[CAR99] S. Card, J. Mackinlay, B. Shneiderman, Information Visualization: Using Vision to Think, Morgan Kaufmann (1999).

[CHI92] R. Chimera, “Value bars: An information visualization and navigation tool for multi-attribute listings,” Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, 293-294 (1992).

[CHI91] R. Chimera, **“**Value Bars: an information visualization and navigation tool for multi-attribute listings and tables”, University of Maryland Department of Computer Science Technical Report CS-TR-2773, http://www.cs.umd.edu/Library/TRs/ (1991)

[HEA97] M. Hearst, “About TileBars”, retrieved from http://www.sims.berkeley.edu/~hearst/tb-overview.html on 6/17/2004. Last modified on August 14, 1997.

[HEA96] M. Hearst, J. Pederson, P. Pirolli, H. Schütze, G. Grefenstette, D. Hukll, “Improving Full-Text Precision on Short Queries using Simple Constraints”, Proceedings of the Fifth Annual Symposium on Document Analysis and Information Retrieval (SDAIR), (1996).

[HEA95] M. Hearst, “TileBars: Visualization of Term Distribution Information in Full Text Information Access”, Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, 59-66 (1995).

[INF02] J. Dürsteler, “TileBars”, InfoVis – The Digital Magazine of InfoVis.net, retrieved May 5, 2004 from http://www.infovis.net/E-zine/2002/num_104.htm. Last modified on October 21, 2002.

[SCH96] B.Schneiderman, “The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization”, Proceedings of the IEEE Symposium on Visual Languages, 336-343 (1996).

[UCB04] University of California, Berkeley. Digital Library Project: TileBar Searching. Retrieved from http://elib.cs.berkeley.edu/tilebars/ on 6/17/2004.