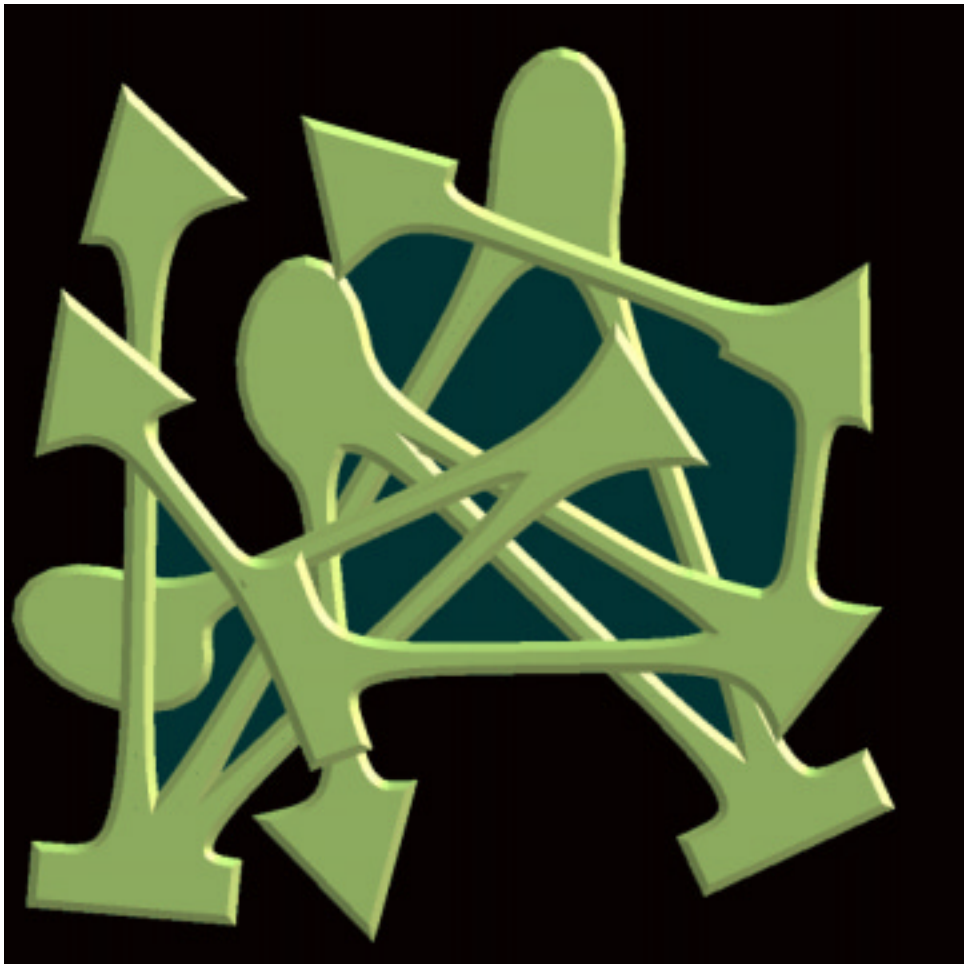
YLEM Journal

Artists Using Science & Technology

number **2** volume **24** January - February **2004**

Visualizing Data Sets Issue 2 of 2



Fillets: soft shapes for easily perceptible connections
(Tuomas J. Lukka, Janne V. Kujala and Marketta Niemelä)

SSN 1057-2031 / (c) 2004 Ylem

Visual Depictions of Search Results: using glyphs and coordinated multiple-views

Jonathan Roberts, Nadia Boukhelifa, Peter Rodgers
University of Kent, UK
J.C.Roberts@kent.ac.uk, N.Boukhelifa@kent.ac.uk, P.J.Rodgers@kent.ac.uk

Searching the web

Searching information on the web is becoming a daily activity for many. The web is growing at a phenomenal rate; some sources say there are currently over 2.5 billion pages available and this number is increasing by 7.3 million pages per day [1,2]. Despite this complexity current search engines do an amazing job. However, search results are mostly presented in rank-ordered lists that the user browses to find interesting information. These traditional representations are spread over multiple pages and the user only views a small proportion of the results in a single window. Certainly the search engines attempt to display the most relevant information in the top 10 ranks but experience shows that we still scroll through many results and load multiple pages.

Textual and Graphical depiction

These text presentations take up a vast amount of the screen real estate. Graphical representations, on the other hand, can display much more information in a smaller space. Perhaps hundreds or even thousands of results could be (abstractly) displayed in the same area as 10 text results. Such abstract depictions would provide a holistic view of the information that the user could refine or zoom into to gather more details from the realization [3]. These graphical representations could provide the user with a synoptic view of many search results (without the need for paging or scrolling), allow clustering operations (portraying similarities between search results), allow the user to quickly find relevant information and then drill-down into specific details. Moreover, such graphical representations could display information-rich content including information/statistics about the sites and pages that the URLs actually address. Conversely, users may expect a textual form as they are used to browsing and manipulating the search results through these interfaces.

In essence, both representation methods (textual and graphical) are valid and have their benefits. Thus, we believe it is useful to simultaneously present the search result information in both textual and graphical forms. Indeed, by coordinating each view together the user would be able to more quickly drill-down to the pages that are most interesting and useful.

Multiple views of search-results

We have developed some visualization tools that display the search result data in multiple linked views, where each window provides a dedicated view on the search results. In this paper, which is a summary of the work given in [4]. We present one system that displays the traditional rank-ordered list of URLs with some additional views that detail statistical quantities of the sites. Moreover, each view is coordinated together such that any enacting user-operation is simultaneously reflected in every view.

Our search-engine retrieves detailed information about the located sites, first by retrieving the traditional URLs and then visiting each site to find this 'rich' information. For example, as well as storing information about the keywords search terms, and retrieving information about the domain-name, URL, text snippets of the pages, we retrieve information such as the size of the HTML file, number of internal and external links (anchors) on a retrieved page, media type (html, text, images, sound), last-modified-date, page structure (e.g. position of the media on the page). We display this extended information in glyph form.

Glyph based depiction of search results

A glyph is a single graphical unit that can portray many variables by adapting its properties. An early example was developed by Chernoff [5] who represented multi-variable data through faces; for example, different values could be represented by changes in the shape of the face, length of the eyes, nose or mouth, the angle of the eyebrows etc. In our designs each search-result is displayed by a single glyph; aspects of the appropriate site are represented by different parts of the glyph. We have investigated different glyph designs; in this paper we present two.

The first glyph design emphasises the domain of the site of each webpage, see Figure 1. Each of the domain names is given a unique symbol (e.g. the addresses: .country, .com, .edu, .org, .net; others are mapped to triangle, circle, square, paralelogram, hexagon, plus-symbol, respectively). The rank is allocated a

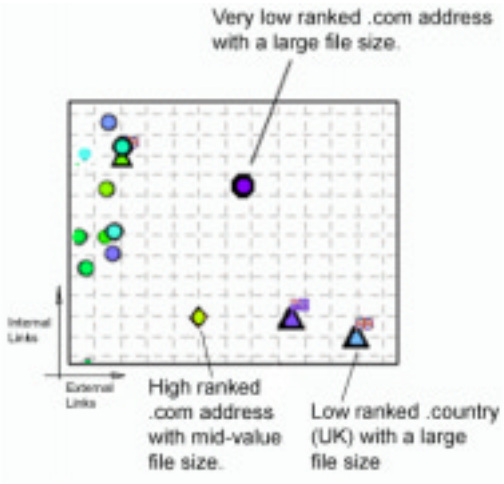


Figure 1. This diagram shows the domain glyph representation.

colour (with a brighter colour representing a higher ranked entity) and the size of the page relates to the border-width of the glyph (a thicker border representing a larger size). The quantity of Anchor tags on the page is mapped to position (the external and internal link quantities are given x and y positions respectively). Moreover, the country of origin – if available – is depicted by a small flag of the country to the right of the glyph. Indeed, glyph placement is important as it allows the user to see clusters of similar results, but as subsequent glyphs may occlude previous glyphs, we jitter the glyph placement by a small random amount in both the x and y coordinates, which allows the user to perceive dense regions of similar elements.

The second glyph design is based on quartiles. It is often beneficial to represent information in quartile ranges, for example teachers often allocate students an attainment class (upper, upper-middle, lower-middle or lower ranges). In our usage we evaluate whether the quantities are smaller than the lower, median or upper quartile. If they are, then segments around the side of a cross are filled in (shown as A,B,C,D in Figure 2). Additional parameters could be represented in further concentric rings. We use this technique to represent the measure of external and internal links of a site (the outer and inner segments, respectively). The colour indicates the relevance like the aforementioned glyph; this is also redundantly visualized as the glyph placement, however, it is possible to reassign these different variables to each part of the glyph.

Coordination and Exploration

In any visualization setup it is important to allow the user to explore and investigate different parameterisations of the visualization. More specifically, it is important to coordinate each of the visualizations together such that any user operation is mimicked in other related views [6]. Such coordination enhances the user operation. In our system, when the user highlights (brushes) the mouse over one window, so the same elements in every other view are simultaneously highlighted. Also, when the user clicks on a glyph or a link (in any window) then the page is loaded in the browser window, see Figure 3.

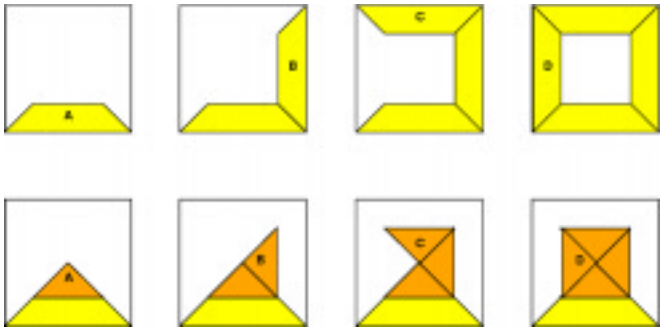


Figure 2. The schematic shows the C-glyph design that allows the observer to quickly realize whether the data is within the lower, middle or upper quartile ranges.

Related Work

The key to success in any such visualization is to develop a suitable mapping to exchange the (abstract) search result data into an appropriate graphical form. Other researchers have developed alternative abstract methods that may be useful in this domain. For example, systems such as SeeSoft [7] and WebTOC [8] reduce the text itself by abstracting lines of text into horizontal graphical lines (a technique called greeking). Some systems use dots or coloured areas to display the search result information such as Dotfire [9] (used for searching digital libraries) or Sparkler (that visualizes session details) [10]. Other systems also realize multiple variables in glyph representations, e.g. Mann [11] uses a technique named TileBars [12] to map the position of the keywords and the lengths of the documents (here tilebar information is laid out adjacent to the text information) and systems such as xFind [13] plot glyphs on an x, y axis representing document size and relevance respectively. Finally, Cugini et al [14] use different presentation styles including spirals and three-dimensional axis designs.

Summary

We encourage the use of information-rich realizations especially in the area of web-search result visualization. To this end we have developed a system that can visualize search-result data using glyphs. In our evaluations, apart from being distracted by the slowness of the data-gathering, the users liked the different visual depictions, and found, in particular, that the C-glyph was good at representing sites that contain lists of links (often known as luminous sites) that are useful for finding general information about a subject.

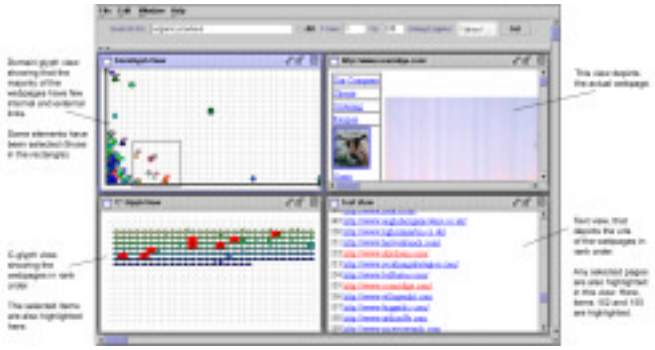


Figure 3. The diagram shows a screen shot of the application, in this example the user searched for "organic produce".

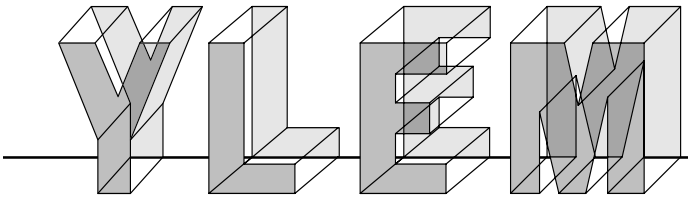
Acknowledgements

This work has been supported by the EPSRC (grant reference: GR/R59502/01). Further information about this project may be found at http://www.cvev.org

References

1. Brian Murry, Alvin Moore. Sizing the Internet, Cyveillance White Paper. http://www.cyveillance.com/web/us/. July 2000.
2. How much information? http://www.sims.berkeley.edu/ research/projects/how-much-info/internet.html, 2002.
3. Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In Proceedings of the IEEE Symposium on Visual Languages, pages 336-343, Washington, September 1996.
4. Jonathan Roberts, Nadia Boukhelifa, and Peter Rodgers. Multiform Glyph Based Search Result Visualization. In Proceeding Information Visualization 2002. IEEE, pp. 549-554. July 2002.
5. H. Chernoff. The Use of Faces to Represent Points in k-Dimensional Space Geographically. Journal of the American Statistical Association, Volume 68, 1973, pages 361-368.
6. Jonathan C. Roberts. On Encouraging Coupled Views for Visualization Exploration. In, Visual Data Exploration and Analysis VI, Proceedings of SPIE, volume 3643, pp. 14 -24. January 1999.
7. S. Eick, Graphically displaying text, Journal of Computational and Graphical Statistics, 3(2):127-142, June 1994.
8. Ben Shneiderman, David Feldman, and Anne Rose WebTOC: A Tool to Visualize and Quantify Web Sites using a Hierarchi-

cal Table of Contents. Technical Report CS-TR-3992, February 1999.
9. Ben Shneiderman, David Feldman, Anne Rose, and Xavier Ferré Grau Visualizing Digital Library Search Results with Categorical and Hierarchical Axes. HCIL Technical Report No. 99-03. June 1993.
10. K. Perrine S. Havre, E. Hetzler and E. Battelle. Interactive Visualization of Multiple Query Results. In Proceedings of the IEEE Symposium on Information Visualization 2001. InfoVis'01, pp 105-112, 2001.
11. Thomas M. Mann, Visualization of WWW-Search Results, DEXA Workshop, pp. 264-268, 1999.
12. Marti A. Hearst. TileBars: Visualization of Term Distribution in Full Text Information Access. Proceedings of CHI'95. May 1995.
13. K.Andrews, C.Gütl, J Moser, V Sabol, W Lackner. Search Result Visualization in xFind. In Proceedings UIDIS. IEEE Computer Society. Pp. 50-58. 2001.
14. J Cugini and S.Laskowski and M.Sebrechts. Design of 3D Visualization of Search Results: Evolution and Evaluation. In Visual Data Exploration and Analysis VII, Proceedings of SPIE Vol 3960. pp 198-210. 2000.



n. pronounced eylum, 1. a Greek word for the exploding mass from which the universe emerged.

An international organization of artists, scientists, authors, curators, educators, and art enthusiasts who explore the Intersection of the arts and sciences. Science and technology are driving forces in the contemporary culture. YLEM members strive to bring the humanizing and unifying forces of art to this arena. YLEM members work in new art media such as Computers, Kinetic Sculpture, Interactive Multimedia, Holograms, Robotics, 3-D Media, Film, and Video.

YLEM
P.O. Box 2590
Alameda, CA
94501
USA

Contact Information

YLEM Main Contact
P.O. Box 2590
Alameda, CA 94501
http://www.ylem.org
ylem@ylem.org

YLEM Membership Issues
members@ylem.org

YLEM Journal Items
Loren Means
149 Evelyn Way
San Francisco, CA 94127 USA
lorenmea@pacbell.net