Visualizing Multiple Websites with Views of Structural Growth, Dynamic Performance and Peer Comparison

Simon Fong, Ho Si Meng, Ma Cheng I
Department of Computer and Information Science
University of Macau
Macau SAR
ccfong@umac.mo, smho@dsej.gov.mo, iris.ma@venetian.com.mo

***Abstract*—Website visualization is an important research topic in data analytics that has its practical applications in performance management reports and navigational problem diagnosis. Researchers from their previous works studied different forms of visualization mainly as 3D linked graphs. In this paper, we propose an innovative tree model that represents website structures as radial and concentric circles of nodes, and users' interactions with the websites as staggered threads. These two types of visualization, when used together, effectively express the growth of a website in a static view and the usage of its services in a dynamic view respectively. Furthermore, the proposed techniques could be placed collectively together for visualizing websites comparatively across different countries or regions as an ecosystem of forest of trees. This paper presents this new concept as well as the features of the custom-made simulator programmed in Java.**

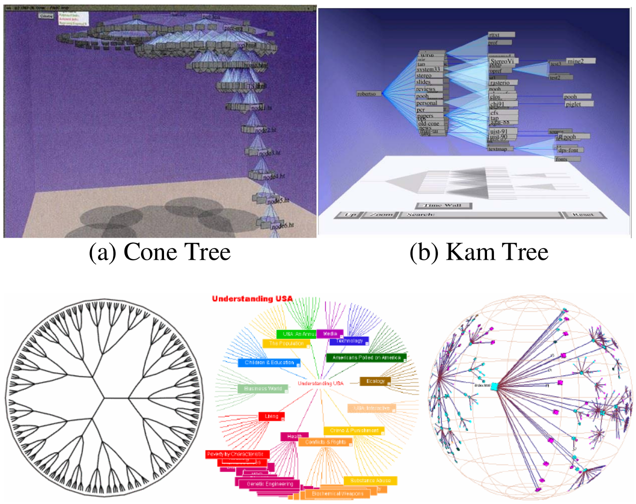
Keywords-Website visualization; Web ecosystem.

I. INTRODUCTION

Website visualization refers to some graphical representation of website information such as the hierarchical structures, the usage data and the natures of the contents etc. The graphical display, which visually conveys a conceivable kind of relations among the data, helps the user to build a mental model about the websites. This mental model is supposedly more effective to be understood about the insights of the website, than the traditional lengthy reports and/or simple displays of tree structures like site-map.

Tree layouts have a long history in visualizing websites as node-link diagrams. Nodes that represent webpages are directly linked with edges that represent the hyperlinks. Web structure visualization in form of tree layouts has been a popular application in evaluating web usability [1]. The tree graph visually displays the possible click-paths through which users can navigate along a website. Another application is to monitor the growth of a website, by visually checking over the document structure and the amount of the contents in a form of tree, ensuring that the growth in all directions is balanced. A balanced growth means information is evenly distributed across different sections and pages of the website for an all-rounded development.

In the pioneer works [2], websites were commonly represented by Cone Tree and Kam Tree, in Figure 1 (a) and (b). Cone Tree has a root node placed at the apex of the cone dangling from it are the child nodes that lined up uniformly. Cone Trees usually are in 3D, so viewing the child nodes on the other side of the baseline requires spinning the tree like a carousal. It is difficult for the child nodes to be labeled in a Cone Tree. To overcome this difficulty Kam Trees were later invented where the root node is on the most left of the graph and the child nodes are stretching across to the right. Focus+context technique was proposed too for making use of inherent perspective projection in the 3D environment. Cone and Kam Trees are known to be ideal for visualizing some medium sized website of around 100 notes, with more of a focus view on a specific piece of information on the tree. Any detail of a tree can be faithfully displayed. They are however not suitable for visualizing an overall website structure because a compact and zoom-out view may be preferred over a potentially very large number of nodes of a website.



(a) Cone Tree (b) Kam Tree

(c) Hyperbolic Browsers

Figure 1. (a) Cone Tree - Courtesy of katsiafas.com; (b) Kam Tree – Courtesy of ifs.tuwien.ac.at; (c) Examples of Hyperbolic Browser [4]

Hyperbolic Browser [3], in a nutshell, maps a tree structure of any size within a bounded circle so that the whole Euclidean plane can always be viewed entirely. In other words, a large number of nodes (hence a large website) can be represented in a single snapshot of hyperbolic display. Relational information among the nodes and links can be shown through coloring or label placement schemes. Hyperbolic Browser still is limited by space constraints.

II. WEBSITE STRUCTURE AS RADIAL TREE PATTERN

In this paper, we extended the hyperbolic tree representation into a radial tree pattern by adding in different colors and sizes of the tree nodes that cluster to a common parent node. The concept derives naturally from the hierarchy of a site-map where any parent node of the tree embraces a cluster of child nodes, and child nodes of the child nodes across several subordinate levels. All these nodes would be visualized collectively by the same color originated from the topmost parent node. In this case, colors represent different major categories of webpages in a website. This is important for website evaluation because the distribution of the colors indicates whether the major categories of website grow in balance or not. Furthermore, the sizes of the nodes symbolize the amounts of information carried by the corresponding web pages. The combination of colors and sizes allows a user to visually check if the website has any biased proportion of information (in terms of weights represented by sizes of nodes) and number of web pages (in terms of spread of colors represented by the numbers of the nodes). Any too heavily biased part of the website can be early identified and the development could be subsequently rectified. Moreover the same radial tree pattern can be used to show the popularity of each section or each page of the website, for the radial tree pattern essentially is representing the whole website structure. Popularity in the context of a website is expressed as the number of hits or page views on each page. Similar to the concept of *heat map*, hot spots of the website can be shown in different colors. When these two modes are used together, managers can evaluate how much resources are invested in various parts of the website contents and how well they are favored by frequent visits.

Figure 2. Radial Tree Patterns; Above – the pattern is showing the static structure of the website with different weights at each categories of pages; Below – the same pattern now uses coloring to show the popularity of each part of the website.

This visualization scheme would be particularly useful for organization or a government that wants to evaluate a collection of their websites vis-à-vis. For example, an e-government portal usually consists of many departmental websites, each of which operates as an individual site that has its own team to update the information and used by different segments of citizens. In this case, multiple websites could be monitored comparatively as a row of radial tree patterns. At one glance, the user can spot the strength or weakness of each site from observing over its structure and weights of contents, as well as popularity. If the patterns are collected over a long time, the growth of these websites can be longitudinally studied and visually compared too.

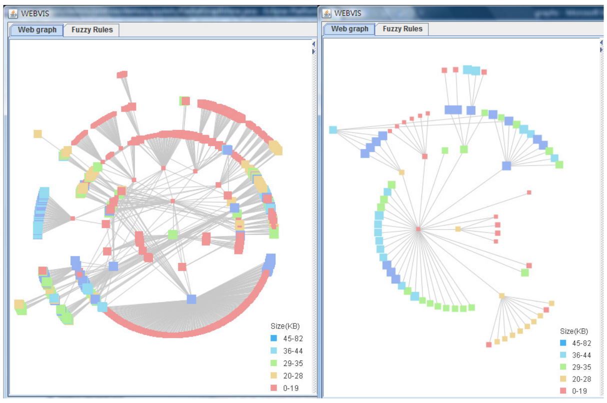


Figure 3. Visual comparison of Radial Tree Patterns of two websites; Left – The website of Education and Youth Affairs Bureau of Macao showing complex and large structure with six levels. It contains more than 900 pages. The node color is indicating the page size. Most pages are in pink having little content.; Right – The website of Fundo de Segurança Social, Macau with simple and small structure having five levels. It contains nearly 100 pages. Few pages but heavy in contents.

The Radial Tree Patterns are technically programmed in Java with a visualization toolkit called Prefuse. A prototype is implemented from our previous project WebVS [5].

III. WEBSITE VISITS AS FUZZY RULES PATTERN

While Radial Tree Patterns are useful for illustrating the full view of a website and the information pertained at different parts of the website, they fall short in visualizing the association of parts of the websites being visited. This association is also known as clickstream in Web mining where a temporal sequence of visits over a series of webpages frequently occurred. Many research papers in the literature [6] have been focusing on this problem. Visualization of such associations is implemented here along with Radial Tree Patterns because dynamic usages of the website that are represented by web visits complement with the growth of website structures and contents as a holistic approach. Importantly, this holistic approach inspires website visualization to a higher level – collective and comparative visualization of websites as a forest of trees. Essentially the information from the growth of website structure and from the dynamic usages of the website serves as input values for modeling parameters of a forest of tree (FOT). A sample of the fuzzy association rules from WebVS is shown in Fig. 4.

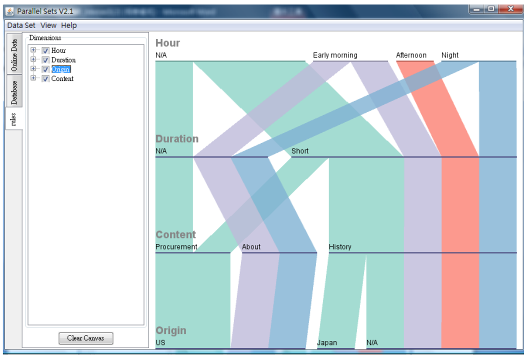


Figure 4. Visualizaton of fuzzy association rules that illustrate which parts of the website are often visited together at different time periods.

As a metaphor, the growth of a website structure is similar to the growth of a biological tree, that spans outwards of their branches and leaves. The longer the establishment of a website in history, the taller and bigger the representative tree is. Metaphorically speaking again, the dynamic usage of a website is analogous to its vital power that promotes the growth the website, hence its representative tree growth. This is based on the assumption that organizations or companies would only invest on the website expansion provided that the website maintains its popularity and thus its worthiness. And the investment is proportional to the rating of popularity of the website measured by hits and page views. Likewise, websites will be left dormant if few people are using them, so are their representative trees withering for a slow death.

IV. FOREST OF TREES FOR WEBSITES COMPARISON

Given the information from static website structures and dynamic usage, which in turn have already been visualized, an innovative modeling scheme of a forest of trees is proposed in this paper. The performance information from a website are mapped to the growth of a tree that represents the website. Table 1 below shows the basic parameters used in modeling FOT. The visualization program is custom developed by using Java and SExI-FS (Spatially Explicit Individual-based Forest Simulator, in full name) from World Agroforestry Center.

TABLE I. MODELING PARAMETERS IN SEXI-FS

SExI-FS parameters	Mapping of web performance indices to the modeling parameters
DBH (Diameter at Breast Height)	Height = Age (history of website in years) DBH = Logarithmic function of (general performance of the website as measured by users feedback/satisfaction, and/or web visits)
Crown Porosity	Amount of information on the website
Survival Probability	Growth = DBH performance. If growth = 0 or below a threshold, survival probability = sustainability of cost for maintaining the website
Mortality Modifier	Economic factors that affect the web operations
Secondary Mortality Probability	Political factors that affect the growth of the websites
Light Sensitivity	Minimum Light Level: The min level for a tree so that the tree can be grown -default 0.15
Imperata Competition Factor	Dynamic usage of the websites in relation to the rival websites

The modeling of the FOT is geographically based, where one tree and its position represent a particular website respectively and the trees are established over different countries or regions. The prime usefulness of visualizing a FOT is for comparing websites of different regions in the perception of different trees in a forest. Tall and strong trees symbolize prosperous websites that are developing well as shown by their well appearance, and they are receiving abundant nutrients for growth by the high usage of the users.

For an example of comparing e-Government websites over an Asia-Pacific region, as shown in Figure 5, the representative trees are growing at their respective countries and they take different shapes and sizes. Some e-Government websites, as presented by their trees are tall (because of long history) but relatively thin in the branches and leaves. That means these websites though have a long history of establishment, their service functions and contents are relatively sparse, which is typical for developing countries. In contrast, some modern cities have their e-Government websites represented by sturdy tree with lush leaves – they are the websites with good functionalities, rich information and popularly used by their citizens.

The FOT is still in progress of being developed, as many technical challenges still imposed. They are basically the detailed mapping of the performance variables from the dynamic usage and association rules of the website. In an abstract level, mapping the structure and simple well-being variable of a tree from the Radial Tree Pattern to a cone tree in FOT is straightforward. However, when the whole modeling is put to function as an ecosystem, many dependent variables exist and they need to be precisely calculated and mapped from the detailed operational variables (KPI's) of a website.

V. CONCLUSION

Website visualization has long been studied by researchers via a variety of visualization models from 3D to hyperbolic browsers. The purpose of website visualization traditionally is limited to allow interactive inspection of navigation paths of the site as represented by nodes and links in the graph. In this paper we proposed a new model that serves slightly different purposes for website visualization. Radial Tree Patterns were used to show the integrity of the website structures, as well as the popular spots of the website by allowing users to visualize the nodes in different colors and sizes. While colors represent homogeneity of categories of web pages, size of the nodes reveal how much contents there are in the webpages. The extent of the balanced growth of a website can be observed from the tree pattern, as well as the distribution of popularity received by the web visitors.

Instead of visualizing a website individually, a novel simulation model is proposed and formulated for comparing multiple websites as a forest; each tree represents a website in a specific region. The trees of the forest are visually compared in terms of their appearances which are derived from performance variables. For future work, we will fine-tune the mappings of the variables from individual trees to FOTs. The forest will be simulated as an ecosystem that depicts competitions and dependencies among the websites.

REFERENCES

[1] Chi Ed., Pirolli P., and Pitkow J., "The Scent of a Site: A System for Analyzing and Predicting Information Scent, Usage, and Usability of a Web Site", Proceedings of the SIGCHI conference on Human factors in computing systems, volume 2, Issue 1, April 2000, pp. 161-168.

[2] Robertson G.G., Card S.K., and Mackinlay J.D., "Information Visualization Using 3D Interactive Animation", Comm. ACM, vol. 36, no. 4, 1993, pp. 57-71.

[3] Lamping J. and Rao R., "Laying Out and Visualizing Large Trees Using a Hyperbolic Space", Proc. 7th Symp. User Interface Software and Technology (UIST 94), ACM Press, New York, 1994.

[4] Neumann P., "Focus+Context Visualization of Relations in Hierarchical Data", Diplom Thesis, Faculty of Computer Science, Universität Magdeburg, Germany, September 6, 2004, pp. 25.

[5] Ho S.M. and Fong S., "Visualizing e-Government Portal and Its Performance in WEBVS", The 5th International Conference on Digital Information Management (ICDIM 2010), July 2010, Thunder Bay, Canada, pp. 315-320.

[6] Baldini P. and Giudici P., "Improving Web Clickstream Analysis: Markov Chains Models and Genmax Algorithms", Mathematical Methods for Knowledge Discovery and Data Mining, IGI, 2008, pp. 233-243.

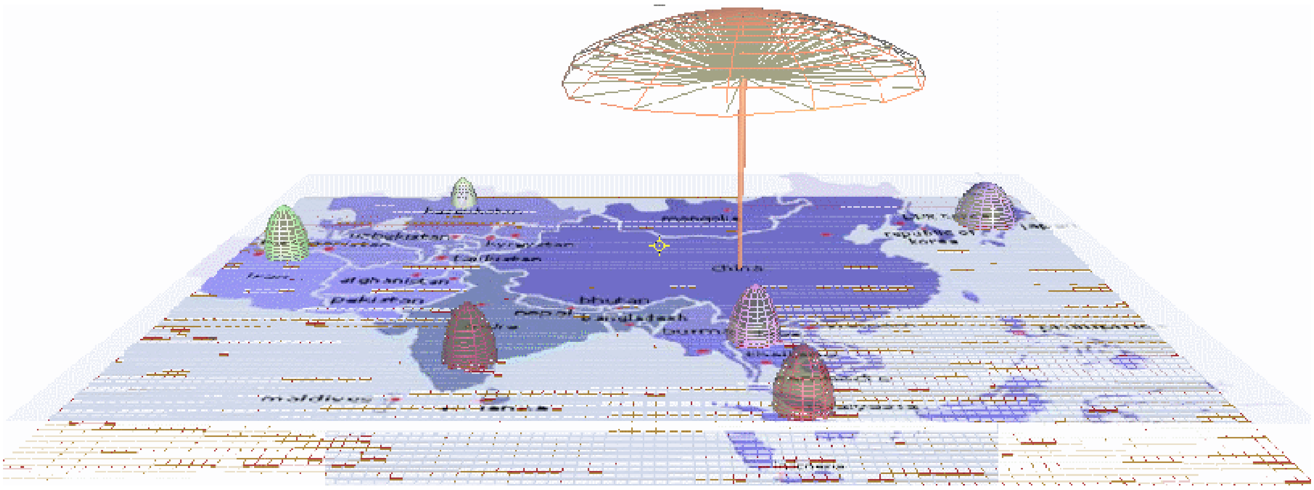


Figure 5. An example of the Forest of Tree visualized by SExI-FS. The graphic shows different e-Government websites are visualized as trees with different shapes and sizes.

Figure 6. Top down view of the Forest of Trees, where the crown of the tree represents the population of the users the website is supposed to serve.

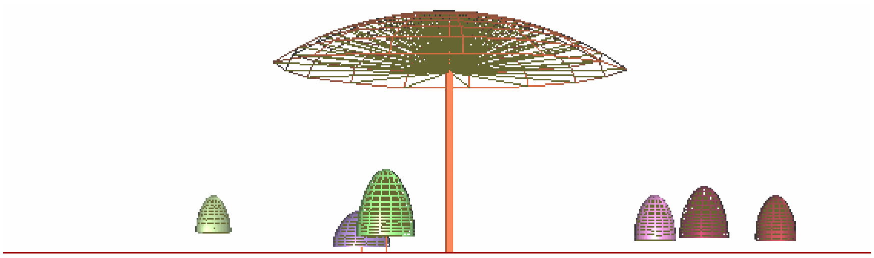


Figure 7. Side view of the Forest of Trees, where the height of the tree trunk represents the age of the website, the sizes of the branches denote the number of services or website categories are available and the density of the leaves represents the amount of information on those webpages.