International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 3, Issue 3, March 2016

Vision Based Analytics and Optimization of Web Pages (V-Lytics)

[1] V.Varsha Rohini [2] P.R.Shreya [3] Dheeksha S [4] E.Nithya
[1][2][3][4] Department of Information Technology
Sri Sairam Engineering College.
[1] varsharohini13@gmail.com [2] shreyaparthsarathyy@gmail.com [3] dheeksha96@gmail.com

Abstract: **Web analytics is the measurement, collection and analysis of webpage data for optimizing web usage and information delivery. However most of web analytics tool limits itself to page statistics and user metrics but does not provide key design inputs for information optimization. This paper aims at extending the scope of web analytics to provide statistics and analysis of each segment of a page, usage of space and the concentration of links in a webpage and its user metrics to help in proper design of the content displayed in a page by using Vision Based Page Segmentation (VIPS) algorithm. The visual block tree which divides a webpage into visual blocks or segments helps us understand the usage of each segment in a page and its content and thereby extending the scope of Web page segment analytics to dynamic web pages and deep web pages. This technique provides us the visibility of the user interaction with the webpage and helps us to place the information in the appropriate segments of the webpage and also effectively manage space in a page and the concentration of links.**

***Index Terms*—Analytics, Design optimisation, vision based technology**

I. INTRODUCTION TO WEB ANALYTICS

Web analytics tools measure the web traffic and report the number of hits, unique users, user metrics, user lifecycle and the like on web pages. These measures indicate only the usage of web pages but not the user interaction. The existing Web design analysis tool limit its input to the generally followed standards. These tools don't give design inputs based on user interaction on the web page.

II. V-DOW AS A SOLUTION

Our proposed solution performs web analytics at a page segment level, analyses the space utilization and uses these information to give design inputs. This solution as two parts:

A. Web Analytics at Page Segment level:

Our solution aims at performing analytics at web page segment level by segmenting a page into visual block tree, at specific depth, using Vision based page segmentation (VIPS) algorithm. VIPS divides the page into hierarchical Visual blocks at a particular depth. The depth, called the Deepest Depth, for a web page at which the leaf node contains only one Distinct Element, is found and Visual Block tree at this depth is obtained. A Distinct Element is a block that contains one image or texts of same font family (Size, font, color, weight and hyperlink). In Fig 2 the red blocks are examples for Distinct Elements; it could be images, text boxes, plain texts, hyperlinks or even buttons.

All the user click actions are clustered from the web page into a Click Cluster. The click cluster might have onClick events, text boxes, images, buttons or hyperlinks. The elements of this cluster are grouped with the leaf nodes of the visual block tree at Deepest Depth. On every click of the user on an element in the Click Cluster, the algorithm first identifies the leaf block that's grouped to it.

The counter for this leaf block is incremented in our database along with the user metrics and the user's click is passed to the respective web page. This tracks the statistics of every leaf node of Visual Block Tree at Deepest Depth. For the statistics of blocks at a higher level in the tree an aggregation can be done on the leaf nodes that belong to it. This gives the page segment analytics.

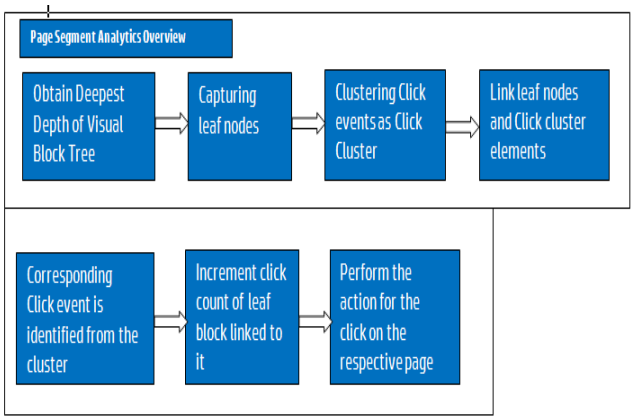

Fig 1: Overview of Web page Segment Analytics

With the insights our solution gives on the segment that's most accessed by the user the important links can be moved to these areas. For example, this could help in placing advertisements on a webpage. This could also be helpful in deep web pages in sorting the results in the result page.

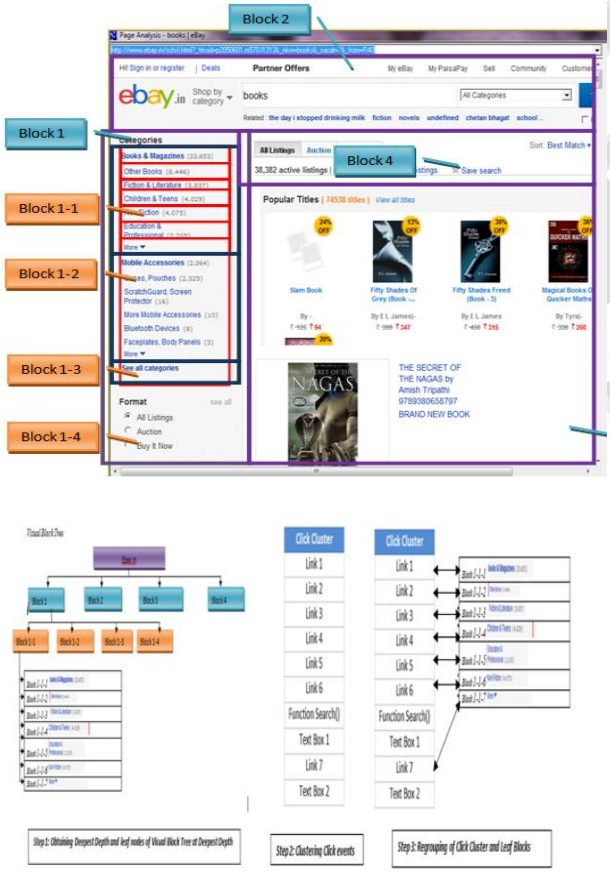

Fig 2. Segmentation using VIPS on sample web page
Fig 3. Overview of Page segment Analytics

B. Analysis of Space Usage:

Analysis of space usage in a webpage can be calculated with the help of Visual Block tree at its Deepest Depth. The algorithm in Fig 4 gives the emptiness percentage of each visual block in a web page. This can be clubbed with the segment statistics to give complete design inputs for a webpage. Our solution can give insights of where most of the links are concentrated in a page and which of these links are ignored by the user. Concentration of the links could be increased in the most visited segment and vice versa. In web forms, the segment statistics and user metrics can help one understand which columns are ignored by users of which region. This can in turn give an interesting inference of what people from a particular region look for or ignore. In Fig 5, an inference could be drawn from the emptiness map that two result items could be placed in a single row. This will help more results to appear in screen without scrolling.

For each Visual Block
IF Root Node
%emptiness of block= Area(sum of all leaf)/area(Page)*100
Else
Depth Difference=Deepest Depth-level of block considered **emptiness%**={[1-(leaf nodes in selected block/Total number of leaf nodes)]*(Depth Dif/Deepest depth)} * 100

Fig 4 Algorithm to Calculate Emptiness
Fig 5 Sample Web Page with Emptiness Marked

III. COMPETITIVE APPROACHES

The existing systems for web analytics focus on the web traffic, the number of hits, unique users, user metrics, user lifecycle and the like on web pages. These measures are calculated only for the web pages but which part of the web page has been used is not covered. There aren't many tools to give design inputs for web pages. The usage of existing systems is restricted to suggesting

guidelines or fixed standards. Our system integrates web page segment analytics and analysis of space utilization of a web page to give design inputs to the user. These inputs include suggestions like where an important content can be places, how the content has to be concentrated to the most used segment of a page or even in ordering the search results in deep web pages.

Current status

Clustering of Click events and counting Clicks on different block are implements and this module is in integration phase. Grouping leaf node and click events is in progress. Emptiness calculation is manually implemented on sample web pages and found to be accurate. Automation of calculation is in progress.

Next steps

This solution can be extended to any web based reporting platform. For example, this can be used on QlikView dashboards to give design inputs.

REFERENCES

[1]VIPS: a Vision-based Page Segmentation Algorithm (2003);Deng Cai , Shipeng Yu , Ji-rong Wen , Wei-ying Ma , Deng Cai , Shipeng Yu , Ji-rong Wen , Wei-ying Ma VIPSDEMO
http://www.cad.zju.edu.cn/home/dengcai/VIPS/VIPS.html