Visualizing e-Government Portal and Its Performance in WEBVS

Ho Si Meng, Simon Fong
Department of Computer and Information Science
University of Macau, Macau SAR
ccfong@umac.mo

Abstract

An e-Government portal is responsible to provide news, information and services to citizens, merchants and tourists in a reliable way. Previously, we proposed a Web Monitoring System (WMS) that monitors the performance of a web portal in real-time [1]. Web portal structure is complex to analyze though it is important to discover useful patterns, evaluate performance, locate anomalies and identify problems. Visualization is helpful in performance evaluation and interpretation of web usage mining results. In this paper we proposed a tool comprised of visualization and fuzzy association rule mining, called Web Visualization System (WEBVS), as an extension to the WMS for monitoring and analyzing performances of e-Government websites.

1. Introduction

In the past decades, the rapid growth of ICT has driven governments to digitalize documents, provide information and implement online services in many countries. On the other hand, such use of ICT has increased pressure on governments to improve performance and achieve "a better government". According to [2], as governments are developing more and more electronic services, they are also coming to the realization that they often do not know what kind of e-government citizens want. ICT tools have provided governments with new ways to provide information and to consult with citizens, but determining the preferences of citizens and businesses with regard to the structure and content of electronic services rests a true challenge. Many people would hard-pressed to be able to articulate their expectations of government, even with a full understanding of the technological possibilities.

The target users of an e-Government portal are usually in three groups: citizens, merchants and tourists. Their needs and preferences can be very different. Taking care of them effectively and reliably is the main objective of an e-Government portal. It is necessary to ensure the information and online services are delivered accurately, effectively and efficiently to these three groups of users, hence meeting their expectations and preferences. In [1], we proposed a WMS to constantly monitor the e-Government portal in real-time. The results of WMS however come only in the forms of statistics and performance reports. In this paper, we propose to use WEBVS to visualize the results of WMS by showing the discovered patterns and performance statistics and superimpose them over the website structure in a social network. Presentation of such information with colors and thickness of nodes and edges is easier to understand, interpret and locate anomalies or interesting phenomenon than lists of statistics and numbers report. We believe such a visualization tool will assist the site administrators and analysts in decision making, maintenance and analysis. In addition to the visual graphs, WEBVS also outputs association rules of the web access data obtained from web logs. The association rules reveal the relations between pages (services) visited and the visitor's browsing behavior. In Section 2, we discuss related web visualization works. Section 3 tells that how our WEBVS relates to the WMS. Our WEBVS web visualization is discussed in Section 4 showing the result graphs and describing the features and functions. Section 5 comes the generation and technique of the Fuzzy Apriori Association Rules we use in WEBVS. Finally we draw a conclusion in Section 6.

2. Related Work

In the past decade, researchers have been exploring visualization methods for understanding the activities that happen over websites. [6]. Recent researches focus on the interactive visualization of the web usage data. However, we have not found any interactive tool or application developed specially for visualizing and analyzing web access and usage in an e-Government portal.

The authors in [3] introduced different 2D and 3D visualization diagrams of particular interest, classifying web pages into two classes of hot (with many hits) and cold (with few hits) ones and illustrating behavior of users. The framework enables flexible selection of mappings between data attributes and visualization dimensions for different diagrams. In the case of large scale portal like an e-Government, however, 3D diagrams will be too dense to interpret.

In [4], the Web Knowledge Visualization and Discovery System (WEBKVDS) is mainly composed of two parts: a) FootPath: for visualizing the web structure with the different data and pattern layers; b) Web Graph Algebra: for manipulating and operating on the web graph objects for visual data mining. The authors presented the idea of layering data and patterns in distinct layers on top of a disktree representation of the web structure, allowing the display of information in context which is more suited for the interpretation of discovered patterns. And with the help of the web graph algebra, the system provides a means for interactive visual web mining.

WebViz system, a tool to visualize both the structure and usage of web sites, is proposed in [5]. The structure of a web segment is rendered as a radial tree, and usage data is extracted and layered as polygonal graphs. By interactively creating and adjusting these layers, a user can develop real time insight into the data. The system shows the idea of interactive visual operators and the idea of a polygon graph as a visual cue. The polygonal graphs, however, are not straight forward enough and it needs time to interpret to discover useful information.

3. Proposed Solution

The analytic results from the four modules in WMS together with the web structure are imported to the WEBVS for visualization. The web structure is rendered as a radial tree and the analytic results will be overlaid on it by some visual cues. Web administrations and analysts can select different data attributes and thresholds to visualize. The web graphs rendered by the WEBVS give the idea of how an e-Gov portal is doing. Such presentation of the portal status is easier to understand and locate anomalies or interesting phenomenon than just lists of numbers in a report. Together with the performance reports generated by the WMS, the site administrations and analysts should be able to know the operating status and performance of the portal at a glance. An overview of WEBVS in relation to WMS is shown in Figure 1.

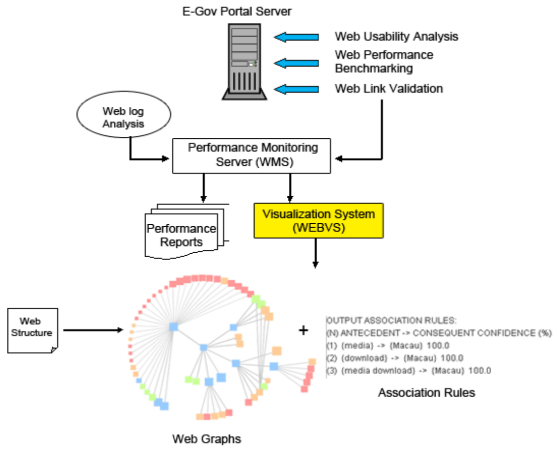

Figure 1. Web-based Visualization System

In figure 2, performance data and web pages information are input to the Integration Engine [3] for data preparation i.e. extracting, cleaning, transforming, integrating data and finally loading into database and later generating XGML for interactive 2D visualization. There are two graph views with analytic results mapping on a web structure radial tree in WEBVS: static graph view visualizes the web page information which only have changes after updates, e.g. the HTML file size, content type, number of words and links in a page, etc.; dynamic graph view visualizes the web performance statistics which is generated by the WMS and changes from time to time, e.g. page views, active time, bounce rate, number of clicks and visitors, etc. WEBVS enables the web administrations and analysts to view both graphs together and select their preferred visual cues for finding interesting patterns. The Fuzzy Sets Generator normalizes the active times on a page for each page view and creates the fuzzy sets for finding association rules.

4. Visualizing e-Government Portal

Definitions of e-government vary in its:

• "The use of information technology to free movement of information to overcome the physical bounds of traditional paper and physical based systems" [7]

• "The use of technology to enhance the access to and delivery of government services to benefit citizens, business partners and employees." [8]

The common theme behind these definitions is that e-government involves the automation or computerization of existing paper-based procedures that will prompt new styles of leadership, new ways of debating and deciding strategies, new ways of transacting business, new ways of listening to citizens and communities, and new ways of organizing and delivering information [9].

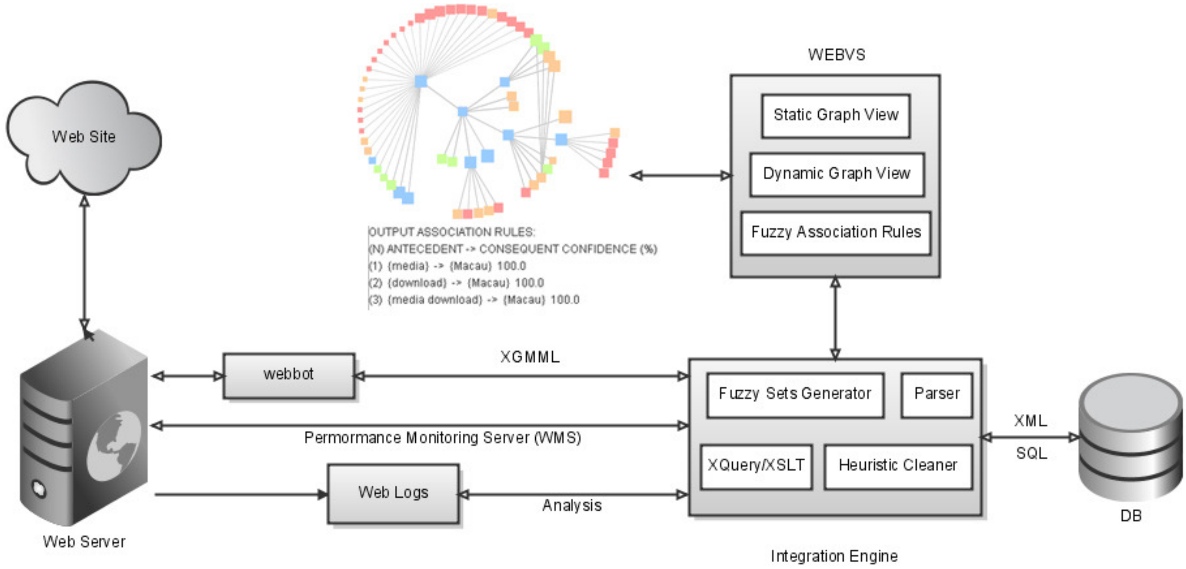

Figure 2. A possible implementation architecture of WEBVS

4.1. Input Data and Assumptions

In addition to the analytic results from the WMS, web structure should be prepared and imported into WEBVS for visualization. Figure 3 shows the input data for the static graph visualization in XGML. Moreover, to discover the preferences and needs of the citizens, merchants and visitors, we have to identify what are their hot pages and what these hot pages are about. Hot pages in [3] are those pages with many hits. The number of hits received by a website is frequently cited to assert its popularity, but this number is extremely misleading. Therefore, our WEBVS measures the hot and cold pages by active time, which is the average amount of time that visitors spend on each page of the site. The web pages are classified into several content groups according to their keywords, for examples: news, information, media, e-services, etc. Then we measure their popularities as hot or cold pages.

```
<graphml xmlns="http://graphml.graphdrawing.org/xmlns">
<graph edgedefault="undirected">

<key id="name" for="node" attr.name="name" attr.type="string"/>
<key id="nSize" for="node" attr.name="nSize" attr.type="double"/>
<key id="words" for="node" attr.name="words"
attr.type="integer"/>
<key id="type" for="node" attr.name="type" attr.type="string"/>
... ...

<node id="1">
<data key="name">index-c.html</data>
<data key="nSize">15</data>
<data key="words">20</data>
<data key="type">info</data>
... ...
 </node>
... ...

<edge source="1" target="2"></edge>
... ...
</graph>
</graphml>
```

Figure 3. Input data for static view in XGML format

Visualizing such access data, our WEBVS can help the web administrations easily identify what content a merchant needs and what content should be translated into a language.

4.2. Web Graphs

In figure 4, WEBVS visualizes the web information in Static view and performance data in Dynamic view. For both static and dynamic views, there is an expandable control panel providing options to configure the mapping between data attributes and visual cues. The static graph view provides the page information in complement to the dynamic web access graph for web administrations and analysts to figure out the anomalies occurring in the web portal in a more efficient way.

The web structure is visualized as radial tree [10]. In the radial tree layout a single node is placed at the center of the display and all the other nodes are laid around it. The entire graph is like a tree rooted at the central node. The central node is referred to as the focus node and all the other nodes are arranged on concentric rings around it. Each node lies on the ring corresponding to its shortest network distance from the focus. Any pair of nodes joined by an edge in the graph is referred to as neighbors. Immediate neighbors of the focus lie on the smallest inner ring, their neighbors lie on the second smallest ring, and so on. Radial tree layout is suitable for dynamically changing graphs since the addition or deletion of a node perturbs its siblings only by a small amount, especially as the graph becomes dense. In WEBVS, the focus node can be switched around by simply clicking at one of the other nodes.

In figure 5, nodes are colored by content type, i.e. green nodes represent pages with media content and red nodes are downloads, while larger nodes indicate that the pages have larger file size meaning the page contains more content. This graph, when compared with the dynamic graph with node size representing active time in dynamic view, can clearly show if visitors spend more time on a web page with more content. If not, the page is having low performance and the web administrator must find out the reason and take action to improve, maybe by adding a link in a related hot page.

Figure 6 is a visualization of web structure as a radial tree showing what the hot pages are in the dynamic graph view. The green nodes are those pages with active time above the threshold set by the site administrator. They are the hot pages on which users spend most time in browsing and interacting actively with the content. The thickness of the edge shows the usage traffic. Site administrator or analyst can select usage data of users from one country or one user group to visualize. The analytic results from the web log analysis module on WMS contribute to this function. Actually, with the results from WMS, WEBVS can visualize various web graphs of particular interest. A site administrator can interactively find out where broken links exist if any, where page load time is higher than an acceptable range, and other phenomena of interest.

There are various indicators of a website that can be shown in the static and dynamic views in our WEBVS with different visual cues. It will have a more readable, meaningful and interesting web graph when quantitative indicator relating to size represents with node size and categorical or unevenly distributed indicator represents with node color. Table 1 shows a typical mapping we found in WEBVS.

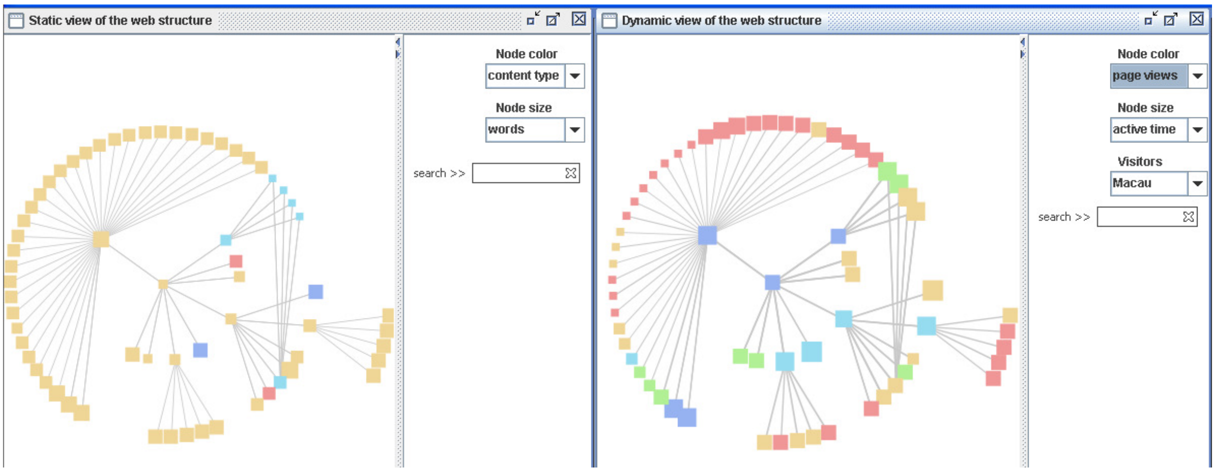

Figure 4. The user interface of WEBVS showing both Static and Dynamic views

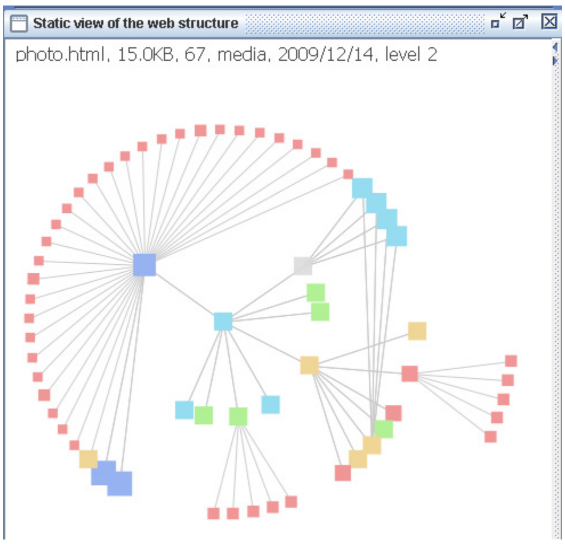

Figure 5. Details are shown when mouse-over a node

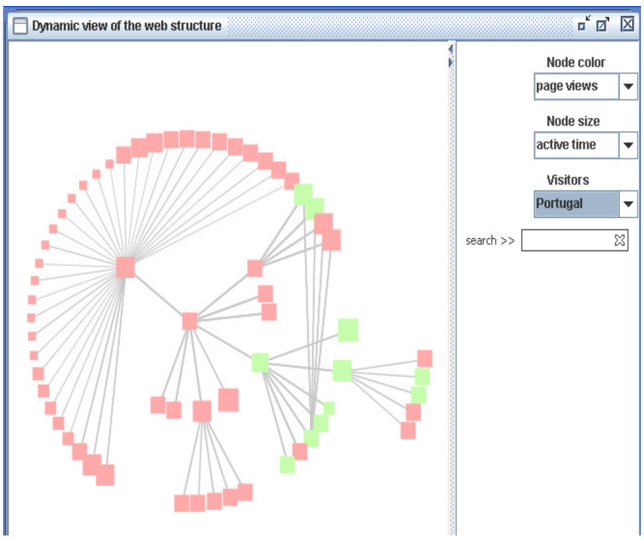

Figure 6. Green nodes are the popular pages interested by users from Portugal

Table 1. Mapping of indicators with visual cues to get meaningful web graph

	Indicators	Node Size	Node Color	Edge Thickness
Static View	File Size	√		
	Number of Words	√		
	Content Type		√	
	Number of Links	√		
	Hierarchy Level		√	
	Last Update		√	
Dynamic View	Page Views	√		
	Active Time	√		
	Bandwidth	√		
	Bounce Rate		√	
	Entry Page		√	
	Exit Page		√	
	Number of Errors		√	
	Number of Clicks			√

5. Fuzzy Apriori Association Rules

Association rule discovery techniques are generally applied to databases of transactions where each transaction consists of a set of items. The goal is to discover all associations and correlations among data items where the presence of one set of items in a transaction implies (with a certain degree of confidence) the presence of other items. In the context of Web usage mining, it is used to discover the correlations among references to various pages available on the server by a given client. Each transaction is comprised of a set of URLs accessed by a visitor [12]. Conventional Association Rule Mining (ARM) algorithms usually deal with

datasets with categorical values and expect any numerical values to be converted to categorical ones using ranges. In real life, data is neither only categorical nor only numerical but a combination of both. And the general method adopted is to convert numerical attributes into categorical attributes using ranges. The problem with the above approach to dividing ranged values into sub-ranges is that the boundaries between the sub-ranges are crisp boundaries. Fuzzy Association Rule Mining (FARM) is intended to address the crisp boundary problem encountered in traditional ARM. The principal idea is that ranged values can belong to more than one sub-ranges is that the boundaries between the sub-ranges are crisp boundaries. Fuzzy Association Rule Mining (FARM) is intended to address the crisp boundary problem encountered in traditional ARM. The principal idea is that ranged values can belong to more than one sub-ranges , we say that the value has a membership degree, μ, that associates it with each available sub-ranges. For example, we define three age categories: Young, Middle-aged and Old, and then ascertain the fuzzy membership (range [0, 1]) of each crisp numerical value in these categories. Thus, Age = 35 may have μ = 0.6 for the fuzzy partition Middle-aged, μ = 0.3 for Young and μ = 0.1 for Old.[11] Thus, by using fuzzy partitions, we preserve the information encapsulated in the numerical attribute, and are also able to convert it to a categorical attribute, albeit a fuzzy one. Therefore, many fuzzy sets can be defined on the domain of each quantitative attribute, and the original dataset is transformed into an extended one with attribute values having fuzzy memberships in the interval [0, 1].

To measure the success or popularity of a page, as we discuss earlier, we use the active time instead of number of hits. When a visitor browses through the internet, he/she will stay longer in pages that have updated, useful and attractive content. In web log files, we can find out the web usage data of a visitor telling us where he/she locates, what pages viewed and the active time spent on a page. We have no problem in defining the locations of the visitors into categories of countries or regions. However, to discover what content they are interested in from a set of pages, we can only measure the length of the active time. In this case, we define the categories as content groups of the pages which are the same as we have defined earlier for the static view web graph, i.e. News, Information, Download, Media and e-Service. Then normalize the active time on a page into the interval [0, 1] by decimal scaling normalization to generate the membership degree of the categorical attribute.

As in classical association rules, $I=\{i_1, i_2, \dots, i_m\}$ represents all the attributes appearing in the transaction database $T=\{i_1, i_2, \dots, i_n\}$. I contains all the possible items of a database, different combinations of those items are called itemsets. Each attribute i_k will associate with several fuzzy sets. In order to represent the fuzzy sets associated with i_k , we use the notion:

$F_{i_k}=\{f^1_{i_k}, f^2_{i_k}, \dots, f^l_{i_k}\}$, where $f^j_{i_k}$ is the j^{th} fuzzy set.

Therefore, we have the attribute Page as follows: F_{Page} = { News, Information, Download, Media, e-Service } and obtain a fuzzy database as in Table 2.

Table 2. Sample fuzzy database

News	Information	Download	Media	e-Service
0.5	0.2	0.3	0	0
0.2	0	0.3	0.5	0
0	0.2	0	0	0.8
0.1	0.2	0.2	0.5	0
0.6	0	0	0.4	0

After setting the support and confident, perform Fuzzy Apriori algorithm and get interesting rules as shown in Figure 7. We can see that visitors who spent longer time on pages with the content of media and download are from Macau.

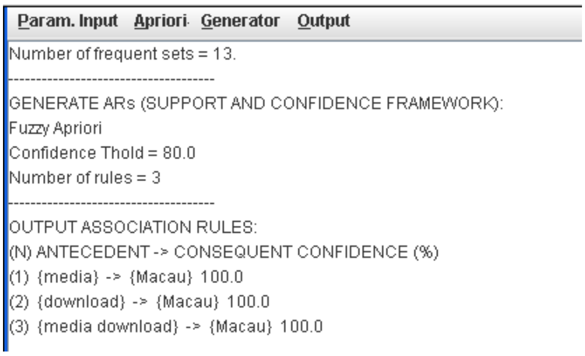

Figure 7. Association rules generated by applying FARM

6. Conclusion

The information and online services of an e-Government portal should be delivered accurately, effectively and efficiently to different groups of users to meet their expectations and preferences. Knowing the needs and preferences of users is very important. For this purpose, WEBVS is proposed to visualize the analytic and performance results of WMS together with web page information in web graphs and generate interesting and meaningful association rules. Such web graphs and association rules help the web administrator to discover and locate where anomalies or interesting phenomenon are and, hence, take actions to improve the portal when necessary, for providing a successful e-Government portal.

7. References

[1] S. Fong, C. S. M. Ho, "A Web-based Performance Monitoring System for e-Government Services", *Proceedings of the 3rd International Conference on Theory and Practice of Electronic Governance*, ACM Press, Bogota, Nov 2009, pp.74-82.

[2] E. Lau, "Challenges for E-Government Development", *In 5th Global Forum on Reinventing Government*, Mexico City, 5 Nov 2003, pp.1-18.

[3] A. H. Youssefi, D. J. Duke, M. J. Zaki, "Visual Web Mining", *Proceedings of the 13th international World Wide Web conference*, New York, May 2004, pp.394-395.

[4] J. Chen, L. Sun, O. R. Zaïane, R. Goebel, "Visualizing and Discovering Web Navigational Patterns", *7th International Workshop on the Web and Databases*, Paris, June 2004, pp.13-18.

[5] J. Chen, T. Zheng, W. Thorne, D. Huntley, O. R. Zaïane, R. Goebel, "Visualizing Web Navigation Data with Polygon Graphs", *Proceedings of the 11th International Conference Information Visualization*, July 2007, Zurich, pp.232-237.

[6] E. H. Chi, "Improving Web Usability through Visualization", *IEEE Internet Computing archive* Volume 6, Issue 2, March 2002, pp.64-71.

[7] http://www.cddc.vt.edu/digitalgov/gov-publications.html, September 2002.

[8] Deloitte, Touche, "At the Dawn of e-Government: The Citizen as Customer", *A global public sector study by Deloitte Consulting and Deloitte & Touche* January 2003, pp.1-36

[9] R. W'O Okot-Uma, "Electronic Governance: Re-inventing Good Governance", Commonwealth Secretariat, London, August 2002, p.5.

[10] Di Battista, G., Eades, P., Tamassia, R., and Tollis, I. G, "Graph Drawing: Algorithms for the Visualization of Graphs". Upper Saddle River, N. J: *Prentice Hall*, 1999.

[11] A. Mangalampalli, V. Pudi, "Fuzzy Logic-based Pre-processing for Fuzzy Association Rule Mining", *IEEE International Conference on Fuzzy Systems*, August 2009

[12] R Cooley, B Mobasher, J Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web", *9th International Conference on Tools with Artificial Intelligence*, 1997, pp.558-567.