

DESIGN QUALITY METRICS FOR A WEB PAGE: A WEB APPLICATION

Shazia Arshad and A. Shah

Dept. of Computer Science & Engineering, University of Engineering and Technology, Lahore

With the growth of World Wide Web (WWW) the number of Web Based Hypermedia Applications (WBHAs) have also increased at a tremendous rate. Issues related to the design of a web page which is a basic unit of WBHAs have forced consideration of different metrics those can lead to an efficient design of a Web page. It is expected that by using these design metric, an efficient design can be produced. Further research can be carried out to accommodate the general web-based applications including the Information retrieval systems.

Keywords: WBHAs, Web Page, Metrics, Web Page Design Metrics

INTRODUCTION

The rate of information on the Web is growing at a tremendous rate. It is estimated that this rate is 300% /annum (Shah and Shoaib, 2005). With this increase in growth rate of information on web, many web based WBHAs have emerged on Web. WWW is a medium which provides an equal opportunity to all large or small organization's products and all other activities related to them. The structure of the WWW consist of large number of interlinked pages, many gateway systems and other design issues. Thus the need of a well structured design and parameters for quality assurance is obvious. There are many approaches for the design of a web application. Some of these include RMM (Isakowitz *et al.*, 1995), HDM (Garzotto *et al.*, 1993), EORM (Lange, 1994) or OOHDM (Schwabe and Rossi, 1995) etc. To consider the metrics for the web base application is primarily the purpose of this research. A web page is a basic unit of any WBHAs. Any web application running on web basically interacts with the users with these basic units. The design of a web page has direct impact on web application. An efficiently designed web page will contribute towards the increased efficiency of the WBHAs. Therefore consideration of the metrics for the web page is the first step towards designing an error free web application. In this paper we have discussed almost all the metrics that have been proposed for a web page. New metrics have been proposed to have an efficient web application design.

Web Application

Web application is an application which is placed over a network (internet or intranet) and is accessed through a web browser. When a web application runs then all or some of its parts are downloaded from the web. This is done through the use of browser or applets. When a web page is retrieved then code is executed on the web server as well as code in HTML is brought into the user's machine. A web application can be of any form e.g. client/ server. A client program

resides on the user's machine which accesses data on the web server rather than using the local network server.

There are three kinds of Web Applications i.e. thin client: A kind of web application where most of the processing and storage occurs on remote server but not on local machine, Rich Internet Application: Web applications where most of the processing is done on the client side and server is accessed through remote messaging mechanism and Rich client: Web applications that do not run in the web browser can be delivered as compiled code like Web Application Solutions.

A web-site structure

A website is a collection of web pages, images, videos and other digital assets that are hosted on a Web server, usually accessible via the Internet or a LAN [Wikipedia web site]. It is a document, typically written in HTML or XHTML format, and may provide navigation to other web pages via hypertext links, that is almost always accessible via HTTP, a protocol that transfers information from the Web server to display in the user's Web browser [Wikipedia web page]. All publicly accessible websites are seen collectively as constituting the "World Wide Web" [Wikipedia web site]. Web pages may consist of files of static text stored within the web server's file system (static web pages), or the web server may construct the (X) HTML for each web page when it is requested by a browser (dynamic web pages) [Wikipedia web page].

Web site structure is just like the blue print of the building. It should not be complicated nor need to be very fancy. The website should be organized in such a manner that visitors easily find what they want. The easier its to use, the longer the users will stay with the website, and more they will see of it.

Good web sites structure can easily grow logically. It will be very easy to add new contents without changing the graphical design of the web site build website for you.

Previously 2 tier models were used for web sites but according to Jeff Offutt now 3 tier or n tier model is used according to the situation because 2-tier models were prone to crackers and there were lot of problems with that model such as security, maintainability and scalability.

The hackers and crackers can easily access to the data files in these models.

But now as the tiers have been increased and well organized so the web application can be scalable and easy to maintain because business and presentation logic is separate. Now it was easy for the programmers to understand the code (Offutt, 2002).

As the different software moved to the separate layer like application server and database server so different applications can run in parallel and help in increasing security, reliability, availability, maintainability and scalability etc.

There are three tiers of web based applications. The client, server, data store tiers and the physical distribution can be borrowed from these tiers. Each tier contains several user interface components. While designing these components (Hejda, 2000) different issues come under consideration that affect the speed when downloading the page and reduce response time when accessing data from database as indicated by (Hull, 2006) that Internet delays, using framework like JAVA, J2EE and ASP.NET that are complex but more efficient than PHP or PERL and web browser have impact on the web site performance.

Looking these issues it is necessary to consider different metrics for a web page. Some of the metrics related to a web page are given below.

Web metrics for a web page

Metrics, as we know, refer to standards of measurement. Therefore, web metrics are standardized ways of measuring something that relates to the Web. Web metrics helps organizations to understand, manage and improve their web systems and hence enhance the quality of their online presence.

The metrics we collected measures attributes of different types of entities. Web is an entity. When we discuss about web, set of metrics collected are organized into five categories: Size metrics, Reusability metrics, Complexity metrics, Effort metrics and Confounding factor. The metrics and corresponding categories are described according to Mendes *et al.*, (2001), Fewster and Mendes (2002) are given as under 1. Size metrics: number of pages belonging to a particular site or domain, or alternatively, the required storage size. Different questions could be asked about the web such as "how large is the Web," "how fast

does the Web grow". The web site metrics with respect to size can be defined as a) Page count: number of html and shtml files used in the web application pages b) Media count: number of media files used in web application c) Program count: number of cgi scripts, JavaScript files, Java applets used in the application d) Total allocation: the total space (Mbytes) allocated for all the html or shtml pages used in the application e) Total media allocation: total space (Mbytes) allocated for all the media files used in the application f). Total code length: total number of lines of code for all the programs used by an application.

2. Reusability metrics are the measurement of web based components that can be used in other web applications or with in a same application and includes a) Reused media count: number of reused/modified media files b) Reused program count: number of reused/modified programs c) Total reused media allocation: total space (Mbytes) allocated for all the reused media files used in the application d) Total reused code length: total number of lines of code for all the programs reused by an application.

3) Complexity metrics are measures of the web application complexity like a) Connectivity (Total number of internal links): It doesn't include dynamically generated links b) Connectivity density²: calculated as *Connectivity* divided by *Page Count* c) Total Page Complexity: calculated using following formula (Mendes *et al.*, 2001).

$$\sum_{1}^{PageCount} \frac{PageComplexity}{PageCount}$$

d) Cyclomatic complexity²: calculated like this (Connectivity - Page Count) + 2 e) Structure: measures how the main structure (backbone) of the application is organized: sequence, hierarchy, or network.

4. Effort metrics are factors involved in calculating the efforts used in developing web site including a) Total effort: Total effort (TE) = structuring effort (SE)+interlinking effort (IE) + inter planning (IP)+inter building (IB)+ link testing effort (LTE)+ media testing effort (MTE). b) Structuring effort (SE): Estimated elapsed time (number of hours) it took a subject to Structure an application c) Interlinking effort (IE): Estimated elapsed time (number of hours) it took a subject to interlink the pages in order to build the application's structure e) Interface planning (IP): Estimated elapsed time (number of hours) it took a subject to plan the Application's interface. f) Interface building (IB): Estimated elapsed time (number of hours) it took a subject to Implement the application's interface g) Link testing effort (LTE): Estimated elapsed

time (number of hours) it took a subject to test all the link on an application h) Media testing effort (MTE): Estimated elapsed time (number of hours) it took a subject to test all the media on an application. 5. Confounding Factors include a) Experience: It measures the authoring/design experience of a subject using a scale from 0 (no experience) to 4 (very good experience) b) Type measures the type of tool used to author/design the web pages.

MATERIALS AND METHODS

Proposed metrics

Web page is a construction unit of a web site. There are two important aspects those have not been considered so far. We have presented these aspects in the form of two new metrics.

Age of web site

Age means "how old a particular resource is". The metric has been introduced to evaluate that how old

match (patterns of data) matches the patterns of data on different web pages and c) *level of similarity* compares the level of similarity with the pre-defined thresh-hold if it is less than thresh hold we can say that the pages on the web sites are less relevant.

RESULTS AND DISCUSSION

To evaluate the proposed metrics we have conducted a case study. This case study has been conducted on the four web pages. The web pages contain data about the "Computer" and the other text is also present. The words on the web pages were calculated & are 50,60,70,80, respectively.

We give the word "Computer" as a query. We found 20, 30, 25 & 15 matching words respectively.

According to relevancy metric we calculate the % of relevancy. We use co-relation by using the values obtained from the web pages. The calculation has been given in table for co-relation value 'r'. Where r is an independent variable & y is a dependent variable.

Table. Calculation of percentage relevancy

x	$(x - \bar{x})$	$(x - \bar{x})^2$	y	$(y - \bar{y})$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
20	-2.5	6.25	40	3.75	14.06	-9.37
30	7.5	56.25	50	13.75	189.06	103.13
25	2.5	6.25	36	-0.25	0.06	-0.63
15	-7.5	56.25	19	-17.25	297.56	129.38
90	0	125	145	0	500.74	222.51

the data is on a particular web site is. The age metric enables the user to decide that either the web site is current or not. This metric helps to check the reliability of the information of web site and ultimately a web page. The metric is calculated as a) *date* defines on which date of the month the web site is modified b) *month* defines on which month of the year the web site is modified c) *year* defines on which year the web site is modified d) *Time* period of modification can be calculated as Time period of modification = (current updating date on web site) – (previously modified date) e) *Freshness of data* can be calculated as Freshness of data = Current date – website updating date.

Relevancy metric

Relevancy metrics measures the relevancy of data on a web site against some query.

The relevancy metric takes into account these parameters to measure the degree of relevancy of any data on the web site such as a) *keywords match* matches the key words on different web pages b) *data*

$$Y = \frac{\text{No of words match}}{\text{Total no of words in document}} * 100$$

$$\bar{x} = 22.5 \quad \bar{y} = 36.25$$

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} = \frac{222.51}{\sqrt{(125)(500.74)}} = \frac{222.51}{250.18} = 0.89$$

(James, 1997)

$$r = 222.51 / \sqrt{(125)(500.74)}$$

$$r = 222.51 / 250.18$$

$$r = 0.89$$

As the value of $r = 0.89$ which is close to '1' so we can evaluate that the web pages in question are relevant to the query and hence all can be retrieved. If any set of web pages have the value less than 0.5 will show that set of web pages in query are not similar to that query and should not be retrieved.

Metrics can be used to find % of relevancy for any set of document present on a web page against some query.

The other defined metrics "Age" has also impact to decide about the authenticity of the data present on any web page. The data, which is present on the web page, needs updation. If page updates are not made it is likely that the data can mislead. The mathematical formula can help the user to find the time the data was put on Web Application Architectures and Development Strategies," IEEE computer, 1997, Pp.322 the web page and from which time it has not been modified. Similar formula has been given to calculate the freshness of the web site.

CONCLUSION

The existing metrics have been studied and reproduced to check their application for web based applications. New metrics have been proposed and validated using the co-relation formula. The results are quite satisfied. New methodology for validation of the metrics has been presented. The research is useful for the web application developers in general. We are actively working to propose the metrics for other web applications including Information Retrieval systems and semantics web etc.

REFERENCES

- Fewster, R. and E. Mendes. 2002. Measurement, Prediction and Risk Analysis for web Application" Proc. IEEE.
- Garzotto, F., L. Mainetti and P. Paolini. 1993. HDM-A Model Based Approach to Hypermedia Application Design. ACM Transactions on Information Systems, Vol .11, No.1, pp. 1-25.
- Hull, M. 2006. Balancing Simplicity and Efficiency in Web Application. Imperial College London / 06.
- Hejda, P. 2000. Architectural Model for User interfaces of Web-based Applications. 25-6.
- Isakowitz, D., E. Stohr and P. Balasubramanian. 1995 RMM: A Methodology for Structured Hypermedia Design. Communication of the ACM, Vol.38, No.8, pp.34-44.
- James, P.K. 1997. Research Design in occupational education. Oklahoma State University
- Lange, D. 1994. An Object-Oriented Design Method for Hypermedia Information System. Conf. Proc. of the 27th Annual Hawaii International Conference on System Sciences, IEEE Computer Society Press, pp.366-375.
- Mendes, E., N. Mosley and S. Counsel. 2001. Web Metrics estimating design and authoring effort. IEE E.
- Offutt, J. 2002. Web Software application quality attribute. Quality Engineering in Software Technology, Nuremberg, Germany pp. 187-198.
- Schwabe B. and G. Rossi. 1995. The Object-Oriented Hypermedia Design Model. Communications of the ACM.
- Shah, A. and M. Shoaib. 2005. Sources of Irrelevancy in Information Retrieval Systems. *International Conference on Software Engineering Research & Practice*, Las Vegas, USA, Pp 877-883.
[http://www.lukew.com/resources/](http://www.lukew.com/resources/WebApplicationSolutions.pdf)
- [WebApplicationSolutions.pdf](http://www.lukew.com/resources/WebApplicationSolutions.pdf)
- <http://wikipedia.org/wiki/web-site>
- <http://wikipedia.org/wiki/web-page>
- <http://www.buildwebsite4u.com/building/structure.shtml>