



A Novel Approach for Mining and Improving Quality of Web Data

G. Pratibha*

Asst. Professor, Matrusri Engineering College
Hyderabad, Telangana, India

Dr. Nagaratna Hegde

Professor, Vasavi College of Engineering
Hyderabad, Telangana, India

Abstract— *The primary goal of website developers is to improve the efficiency of websites use by taking less time to accomplish a particular task. To reach the goal, design should be done in a generalized way reaching the psychology and physiology of users mind. This paper investigates the importance of website design complexity and usability metrics against four frequently visiting web pages of Osmania University in two versions and made a certain comparisons to identify whether the redesigned site meets the usability metrics and how far the design and usability complexity has been reduced.*

Keywords— *Design Complexity, Osmania University, Usability Metrics, Usability Complexity, Website Design Complexity.*

I. INTRODUCTION

World Wide Web (WWW) is a hub of huge number of documents. Web consists of some millions of documents providing useful information to the web surfer. The user who really in need of information currently choosing web rather than physical libraries because web provides most relevant information based upon the user query within less time and with many supporting documents in different formats like images, videos etc. It is the responsibility of the organizations that are providing information, to design their websites as per user requirements and expectations. The primary notion of usability is that a website designed with a generalized user's psychology and physiology in mind must be more efficient, easier to learn and satisfy user needs.

The market is saturating with lot of competing brands in everyday life has made usability more popular in recent years, as organizations see the benefits of researching and developing their products with user-oriented methods instead of methods based on technology orientation.

Usability describes the quality of user experience across websites, software, products, and environments. There is no consensus about the relation of the human factors and usability. Some fix usability as the software specialization with ergonomics focusing on physiological matters and usability focusing on psychological matters.

In chapter II, we studied some literatures and reviewed some methods to be success in developing user expected website. In chapter III, we identified and studied certain design principles and evaluation methods. In chapter IV, we compared 04 most frequently visiting web pages of Osmania University website and mined some metric to evaluate usability and complexity. In chapter V, we identified some design metric values which suggests making few changes to meet the optimization.

II. BACKGROUND

This section, we reviewed some previous works in website usability evaluation area, including several studies that specifically looked at academic websites. Borges et al. (1996) formed general Web pages design guidelines through a study of university and college Web sites. Corry et.al.(1997) conducted a usability evaluation of an existing university Web site. After a need analysis was used to restructure the information contained in the current Web site, a prototype was developed and tested against the existing site. Usability was based on the ability of end user subjects like students, parents, and faculty to quickly and accurately locate answers to a set of evaluation questions. To evaluate they tested the original home page of three sites against a version of the revised home page according to their proposed guidelines. Usability was measured in terms of some metrics by task times on the original and revised home pages. A set of usability principles was created by user interface experts and systems developers to evaluate a Web site prototype. The heuristics included consistency of pages, aesthetic design, and navigational feedback.

Kantner and Rosenbaum (1997) gave brief descriptions of several usability studies that they have conducted and recommended an iterative sequence of heuristics evaluation to find obvious problems, followed by user testing to uncover deeper ones. Their testing seemed to focus on collecting data on task completion time, user made errors, relevant information satisfaction, preferences, and the path taken through the Web site. Morkes and Nielsen (1998) evaluated two versions of a given website that differ in terms of their conciseness, scan ability, and objectivity. The usability of the two sites were compared based on task time, errors, memory recognition and recall, and subjective satisfaction.

Misic and Johnson (1999) applied functional issues to the evaluation of a university Web site against similar sites. The other sites in the study included not only schools but also schools linked to the organizations. The sites were evaluated based on absolute usability and complexity metrics such as speed, ease of finding information. The sites were

evaluated by a group of students according to navigational, content and style, and content information metrics issues' easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

III. DESIGN PROCESS AND EVALUATION

There is several usability and design related issue that require careful validation when designing and using Web sites. The most important of them includes 'up-front' issues such as fixing clear and concise goals for a Web site, determining the content to be placed based on the user requirements, ensuring that the Web site meets the expectations of the user.

The present research suggests that the best way to begin the construction of a Web site is to have many different people propose design solutions in parallel to design and requires conducting the appropriate usability complexity, design tests and after evaluation make the findings to changes the Web site.

In this section we presented some of the identified guidelines to be followed for designing and evaluating a web site.

1. Content placed must be of user expectations.
2. Design the website as per user requirements.
3. Make sure that the web site format meet the user expectations especially in content, navigation, download time and relevant information.
4. Gather user's opinions in design activity to improve completeness and accuracy.
5. Identify the primary goals of the web site design.
6. Focus on achieving a higher range of user performance expectations based on the look and feel of the web site.
7. Consider as many user interfaces as possible.

Set design and usability goals including success rate, time taken to complete in finding specific information and how far the content relevancy is satisfied by the user.

IV. COMPARISON OF FREQUENTLY VISITING WEB PAGES

In this section we had a to look on most frequently visiting web pages of two different sites in terms of certain complexity metrics. We have been considered o4 web pages such web pages are Examination Time Tables, Syllabus, Almanac and Distance Education of Osmania University website and examined the complexity metrics on two versions.

Table I Examination Time Tables Web Page Complexity Metrics

Complexity Metric	Redesigned site	Previous site
Number of graphics	171	6
Graphics size total (in KB)	194.27	150.01
Number of words	475	264
Internal links	113	28
External links	6	0
Same page links	0	0
Page height	768	768
Page width	1024	1024

Table II Syllabus Web Page Complexity Metrics

Complexity Metric	Redesigned site	Previous site
Number of graphics	212	18
Graphics size total (in KB)	351.04	262.31
Number of words	791	138
Internal links	198	20
External links	05	0
Same page links	0	0
Page height	768	768
Page width	1024	1024

Table III University Library Web Page Complexity Metrics

Complexity Metric	Redesigned site	Previous site
Number of graphics	141	56
Graphics size total (in KB)	300.88	42.59

Number of words	374	398
Internal links	124	15
External links	06	0
Same page links	0	01
Page height	768	768
Page width	1024	1024

Table IV Distance Education Web Page Complexity Metrics

Complexity Metric	Redesigned site	Previous site
Number of graphics	138	01
Graphics size total (in KB)	314.83	210.16
Number of words	423	135
Internal links	123	33
External links	07	01
Same page links	0	0
Page height	768	768
Page width	1024	1024

Table V Cumulative 04 Web Pages Complexity Metrics

Complexity Metric	Redesigned site	Previous site
Number of graphics	608	81
Graphics size total (in KB)	1161.02	665.07
Number of words	2063	935

In Table 5, we never considered certain fields like internal links, external links and screen resolution because they never show impact on the download time of the page. We eliminated same page links field as it is 0 for all the pages in the previous and re-defined site.

Generally, a website with most number of graphics and content takes more time to download. Even though the re-designed site is having more number of graphics, content and occupying much more space does not shown impact on the download time. The difference in download times of the previous site and the re-designed one is negotiable.

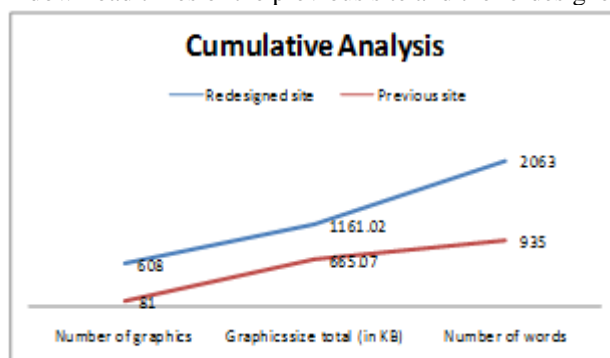


Fig 1: Chart showing comparative analytical values

However these complexity metric values are considered for further experiments, and a better performance of re-designed site.

V. QUALITY EVALUATION OF REDESIGNED WEBSITE

The success of web developers depends upon the delivery of high quality website. Quality is one of the key factors in the growth and success of a product in the market. In recent years, quality of website and its quality in service have become principles. Quality is an ambiguous word which means the degree to which a website meets customer or user needs within short time. The simplest definition of quality is in the mean of customer satisfaction. Robert Glass, summarize the customer satisfaction in mathematical equation as –

$$\text{Customer Satisfaction} = \text{Compliant Product} + \text{Good Quality} + \text{High Performance}$$

Quality assurance includes all the activity-related process to achieve the quality. It is involved from the start of a project. In other words we can say that it is an umbrella activity which is applied on each step in the web development process which controls the insight and oversight quality.



Fig 2: Factors impact on website quality assurance

Therefore quality, reliability, accessibility, usability, adaptability and functionality have become important factors and very crucial for the Web applications.

During the evaluation of website design of Osmania University, certain new issues have been identified which shows somewhat poor quality of design. We listed some of the main factors which show poor quality of website development.

1. Out of 1139 URLs, 285 URLs are identified as broken links.
2. Out of scanned 998 pages, 48 pages were not found in the site map generated.

Table VI Errors Identified In Re-Designed Website

S.No	Error Tag Name	No. of Errors
1	Image	116
2	Bold	24
3	Break	21
4	Paragraph	08
5	Marquee	07
6	Table	03
7	Script	03
8	Style	02
9	Others	08

The figure depicts the percentage of errors identified varied in terms of tag types.

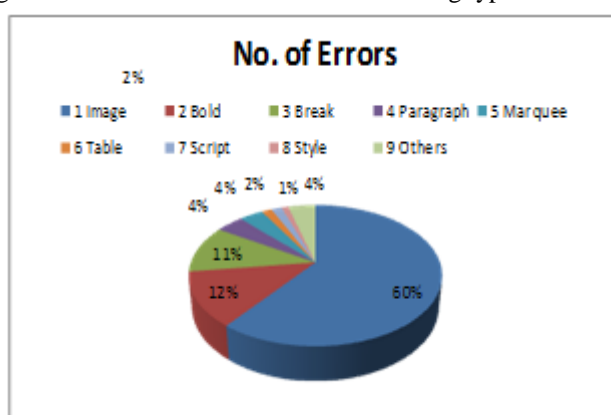


Fig 3: Issues identified in re-defined website in %

The errors or issues highlight missing or conflicting information which caused the validator to perform some guesswork prior to validation, or other things affecting the output. If the guess or fallback is incorrect, it could make validation results entirely incoherent. It is highly recommended to check these potential issues, and if necessary, fix them and re-validate the website.

VI. FUTURE ENHANCEMENTS

Website complexity and design testing includes a range of test and evaluation methods. The testing methods have been categorized as automated, inspection, operational and human performance. Analytical and Statistical based evaluation is needed to judge the quality of a website after completing its design and before moving to onsite. To do so, we have to assign the users to perform a variety of tasks with an operational system while observers note what each user does and record the mined data. Based upon the complexity and design metric values we get, modifications and amendments need to be done.

In future we are planning to perform a major evaluation process on various educational websites and make certain suggestions to the developers so that every designed website will be an error free one.

VII. CONCLUSION

This paper has made contributions in the area of Web site evaluation in terms of design and usable complexity. Our research stated that a website which is good enough in reducing user complexity while in search of their needed data can have issues in website design. The design complexities shown an impact on the performance of the site loosing scope in creating a Web site that all people can use effectively.

ACKNOWLEDGEMENT

I want to take this opportunity to thank all the people especially my guide Dr. Nagaratna Hegde who helped me during my conference sojourn. I understand that it is rather late to acknowledge their contributions, but as the saying goes, better late than never!

REFERENCES

- [1] Nagaraju Mamillapally, (2013): A Comparative Study of Redesigned Website Based on Complexity Metrics, *International Journal of Computer Engineering and Technology(IJCET)*, Volume 4, Issue 3, May-June (2013), pp. 353-358.
- [2] Nagaraju Mamillapally. (2013). A Preliminary Investigation into Complexity and Usability Metrics of a Website. *Proceedings of the Second International Conference on Competency Building Strategies in Business and Technology for Sustainable Development 22nd February 2013.*
- [3] Amitay, E. (1998). Using common hypertext links to identify the best phrasal description of target Web documents. *Proceedings of the ACM SIGIR'98 Post- Conference Workshop on Hypertext Information Retrieval for the Web.* Retrieved February 20, 2003.
- [4] Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A., & Raghavan, S. (2001). Searching the Web. *ACM Pansactions on Internet Technology, 1(1)*, pp.no 2-43.
- [5] Armstrong, R., Freitag, D., Joachims, T., & Mitchell, T. (1995). Web watcher: A learning apprentice for the World Wide Web. *Proceedings of the AAAI-95 Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, pp.no 6-12.
- [6] Borges, J. A., Morales, I. & Rodriguez, N. J. (1996). Guidelines for Designing Usable World Wide Web Pages. *Proceedings of the CHI '96 Conference Companion on Human Factors in Computing Systems: Common Ground*, pp.no 277-278.
- [7] Chi, E. H., Pirolli, P. & Pitkow, J. (2000). The Scent of a Site: A System For Analyzing and Predicting Information Scent, Usage, and Usability of a Web Site. *Proceedings of the CHI 2000 Conference on Human Factors in Computing Systems*, pp.no 161-168.
- [8] Comber, T. & Maltby, J. (1996). Investigating Layout Complexity, in J. Vanderdonckt (Ed.), *Computer-Aided Design of User Interfaces*, pp.209-227.