

## ANALYZING INDIAN WEBSITES

R.B. Kulkarni<sup>1</sup>, Pooja R. Rajeshwarkar<sup>2</sup> & S.K Dixit<sup>3</sup>

---

The use of websites has become rampant now-a-days. But their development procedure is poor. This leads to the need of analysis of websites in order to improve them in every aspect. It may be its efficiency while loading, its look & feel in the browser, its ability to be visible to the search engine, and many more as concerned. The components present in the webpage decide this efficiency. Therefore we will take precautionary steps while developing the website and analyze them after they are developed. Thus, analysis is important to cross-check our ideas. In this paper we will primarily keep Indian websites as our target. We will measure various metrics of these websites and compare them with those of the non-Indian websites.

---

### 1. INTRODUCTION

Analysis means an investigation of the component parts of a whole and their relations in making up the whole. Analogous to this we can define website analysis as, an inspection of web page source (component parts) of a website (whole) and their relations in making up the website.

While surfing the Internet we are often asked to register to view or download any stuff. It frustrates the person using your website. The web allows you to publish your business or content to a very large audience, very easily. But that does not change the need to make sure you understand your audience. By connecting with your visitors, you make them feel more comfortable. If your site "thinks" the way they do, it becomes natural and they will recommend it to friends. The better your visitors feel about using your site, the better your bottom line will be. In addition to design (or behavior) analysis, web analytics can be used to diagnose server or site problems, and measure the effectiveness of marketing and advertising campaigns.[5]

The related work in 2<sup>nd</sup> section is the motivation to write this paper. In 3<sup>rd</sup> section we have presented the statistics for various metrics of Indian websites that are compared to those of the non-Indian websites. Finally we conclude in 4<sup>th</sup> section.

### 2. RELATED WORK

Analysis of websites is mostly done by observing the usage patterns in server logs. It provides the traffic related data and user statistics, which the evaluator must evaluate to predict the problems of website.[1] This is the indirect way of predicting the fault in the website. Caching is the main concern for accuracy of the server log analysis. Besides there is always a need of in-house expert for this

analysis. Instead the direct way would be to analyze the page-level metrics not the log files. Because this approach will take into consideration the metrics like page load time, page rank, etc which are directly related to the behavior of the users. For example if the page load time of a site is more, we can easily say why the users bounce off from the site. This is what the user behavior that we will record in the log analysis approach and then we need to come a long way to conclude this.

However this was understood earlier by many researchers who have progressed their research in this way. Rashmi R. Sinha and two of her colleagues at UC Berkeley[1] have devised a tool to compute 11 quantitative web page attributes. These attributes or metrics are purely related to HTML.

Melody Y. Ivory and Marti A. Hearst at UC Berkeley have also built a tool analogous to grammar checker, called 'quality checker' that is based on empirically derived measures computed over thousands of Web pages. They converted these measures, which characterize the informational, navigational, and graphical aspects of a Website, into profiles for a variety of site types. [1]

This paper may be considered as the future work in this field. All the earlier papers find the metrics and build tools for calculation of them. But, this paper will use some of the tools to actually find the statistics of the sites and compare them. Especially this paper will compare the Indian websites with the corresponding non-Indian websites and evaluate some points after the comparison study. Thus, the work will be beneficial for the Indian website designers to focus on these metrics that will help them to improve their website usage and hence, business.

### 3. EVALUATION & CASE STUDY

In this section, we provide comparison of renowned Indian websites and the corresponding foreign websites. These figures are found by using more than one web analyzers which are listed in the References section. [a,b,c,d]

<sup>1,2</sup>Department of Computer Science & Engineering, Solapur University, Solapur, INDIA

<sup>3</sup>Department of Electronics & Telecommunication, Solapur University, INDIA

E-mail: <sup>1</sup>raj\_joy@yahoo.com, <sup>2</sup>pooja\_rajeshwarkar@yahoo.co.in

### 3.1 Analysis Through Metrics

Here we will consider some metrics termed as Page tagging metrics for analysis. The metrics stated here are the important ones to analyze any website. They are easily understandable by a web developer, so that he can draw some conclusions after looking at the statistics of that metric. These metrics are technically important as well as the look after the important factor of visibility to search engines. Search engine will consider your web page only if it is optimized accordingly. This is called Search Engine Optimization which is a different perspective of analyzing websites. These metrics will thus serve for both the purpose.

By studying tables 1 & 2 (All the statistics of the websites present in both the tables, are taken in march 2011 which may change due to changes in web page design) we can evaluate that the following

1. Page load time: average page load time for Indian websites is 4.62786711 seconds and that of non-Indian websites is 2.36928731 seconds which is half of that of Indian websites. Having slow loading pages can affect website rankings, and ultimately its traffic. Even the traffic may get bounced at a higher rate on slow loading pages. (b)
2. Google PagePank: average page rank of Indian websites is 4.23 and that of non-Indian websites is 6.34. Google uses PR to determine a page's relevance or importance, trust and authority of any website. Domains with a higher PR are likely to rank for more terms than sites with low PR.
3. Inbound links to site: Inbound links send visitors to your web site so more inbound links are required. We find that this number for Indian websites range in thousands while that in non-Indian websites range in lakhs.
4. Broken links: Maximum broken links in Indian websites is 2 while that in non-Indian websites it is 1 (Exception: www3.imperial.ac.in has 23 broken links). Broken links have negative effects on your search engine rankings so it is quite reasonable to be proactive in avoiding them to improve exposure and increase site traffic.

5. 404 error handling: All Indian websites handle 404 errors properly, whereas we find some non-Indian websites have not handled it properly. Handling 404 errors improperly can actually lead to a lot of duplicate content and unnecessarily forces the search engines to rank the page that is not in existence yet.
6. Markup validity: It is very surprising that all the websites fail in this validation. Browsers can accept Web pages and try to display them even if they are not legal HTML. The problem is that different browsers may get hopelessly confused and produce a mangled mess, or even crash. Website development should be done conservatively to avoid this mess. Markup validity thus helps to check if the code is legal.
7. H1 tag: proper writing of the H1 tag have had more stable rankings, as compared to sites that lacked that important element. Search engine crawlers look for your H1 description to determine the relevance of your page. Majority of Indian as well as non-Indian websites do not have H1 tag. May be due to this, the page rank is low for both of them.
8. XML sitemap: Very few Indian and non-Indian websites lack XML sitemaps. An XML sitemap takes a website sitemap and places it into an XML format so it can be easily read by other sources. The main purpose or consumer of an xml sitemap are the major search engines. An XML Sitemap gives you more control of how search engines find and see your website; it lets you present them with a map of all the pages in your website. This makes task of search engines easier to locate your web pages.
9. Duplicate content: Indian websites are badly affected as they have duplicate content because is one of the most common issues that can plague websites from not being found in search engines. Only a few non-Indian websites are having the problem of duplicate content.

Table 1  
Statistics for Indian Websites

Site	Page load time (Seconds)	Google page rank	Inbound links	Broken links	404 error handling	Markup validity	h1 tag	XML sitemap	Duplicate content
www.flykingfisher.com	7.495471	7	52,798	1	Proper	Error in XHTML	Fly Kingfisher	Present	Not found
www.hotelgreenpark.com	2.755446	4	550	1	Proper	Error in HTML	No tag	Present	Found

Table Contd...

Table 1 Contd...

www.10ka20.com	0.37767	4	3,183	2	Proper	Error in HTML	No tag	Not present	Found
www.kesari.in	0.936031	0	0	0	proper	Error in HTML	No tag	Present	Found
pantaloon.futurebazaar.com	4.466624	0	1	2	Proper	Error in HTML	No tag	Not present	Found
www.cafecoffeeday.com	6.67711	5	1,763	1	Proper	Error in XHTML	Found	Present	Found
www.witsolapur.org	1.058301	4	318	2	Proper	Error in HTML	No tag	Not present	Found
www.ignou.ac.in	4.01075	7	45,147	1	Proper	Error in XHTML	No tag	Present	Found
www.tatamemorial.com	13.873401	7	2,945	0	Proper	Error in HTML	No tag	Present	Found

Table 2  
Statistics for Non Indian Websites

Site	Page load time (Seconds)	Google page rank	Inbound links	Broken links	404 error handling	Markup validity	h1 tag	XML sitemap	Duplicate content
www.usairways.com	1.66321	7	328,726	0	Not Proper	Error in XHTML	News and updates	Present	Found
www.grandnewyork.hyatt.com	0.834214	2	12	0	Proper	Error in XHTML	No tag	Present	Not Found
www.hollywood.com	2.04479	7	893,240	0	Not proper	Error in XHTML	No tag	Not present	Not Found
www.thomascook.com	2.662054	6	113,430	0	Not proper	Error in HTML	No tag	Present	Not Found
www.levistrauss.com	4.466624	6	36,193	1	Proper	Error in XHTML	No tag	Present	Found
www.caffenero.com	1.468481	5	3,601	0	Proper	Error in HTML	No tag	Not present	Found
www3.imperial.ac.uk	1.413905	9	125,564	23	Proper	Error in XHTML	Imperial College London	Present	Found
www.ox.ac.uk	1.199102	9	470,103	0	Proper	Error in XHTML	Homepage	Present	Found
www.bumrungrad.com	3.201917	6	48,743	0	Proper	Error in XHTML	No tag	Present	Not found

### 3.2 YSlow Analysis

Now here is the evaluation of websites with a different perspective. We have used a firefox add-on, YSlow, to analyze some factors related to source code of the web page and arrangement of components in the source code. Given below are the pie-charts for home pages of www.flykingfisher.com & www.witsolapur.org [a] (Fig. 1(a) & Fig. 1(b)) which are to be compared with their corresponding non-Indian websites, www.usairways.com & www3.imperial.ac.uk [a] (Fig. 2(a) & Fig. 2(b)). We may

wonder why there is so much difference in their page load times. The main reason is the number of HTTP requests that a browser has to make to load the web page completely. Moreover, Home page of www.flykingfisher.com has total 12 external Javascript, 6 external stylesheets and 19 external background images which are to be optimized. Decreasing the number of components on a page reduces the number of HTTP requests required to render the page, resulting in faster page loads. Some ways to reduce the number of components include: combine files, combine multiple scripts into one script, combine multiple CSS files into one style sheet, and

use CSS Sprites and image maps. Home page of www.usairways.com follows some of the optimizing techniques and we find that it has 12 external javascript files, 4 stylesheets and only 2 external images which reduce the number of HTTP requests to 19. Similarly pie charts for www.witsolapur.org & www3.imperial.ac.uk are compared [a]. As we see more external elements in www3.imperial.ac.uk, we see that there are more HTTP requests. The last two pie charts are for www.usairways.com & www3.imperial.ac.uk[a] (Fig. 3(a) & Fig. 3(b)) when the sites were modified. This proves that, modifications in the web page leads to modifications in the number of HTTP requests required for loading that page. We see here that the modifications done are not favorable because the number of HTTP requests has risen.

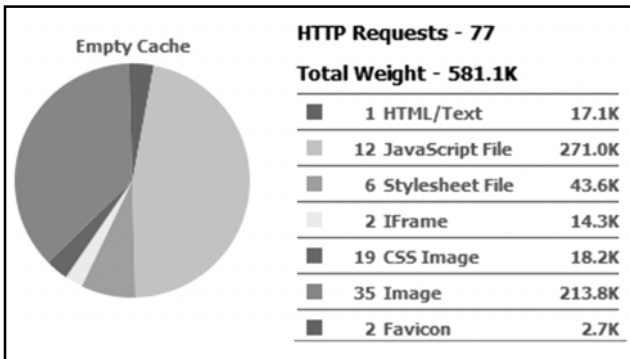


Fig. 1:(a) www.flyingfisher.com

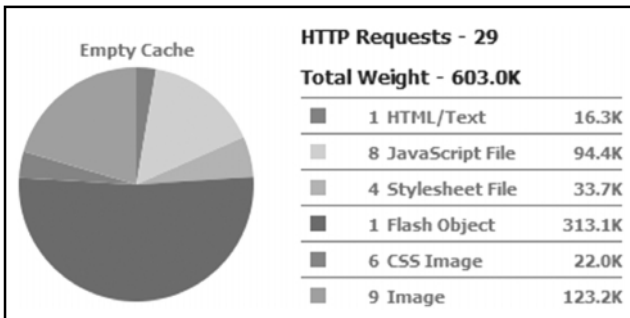


Fig.1(b) www.witsolapur.org

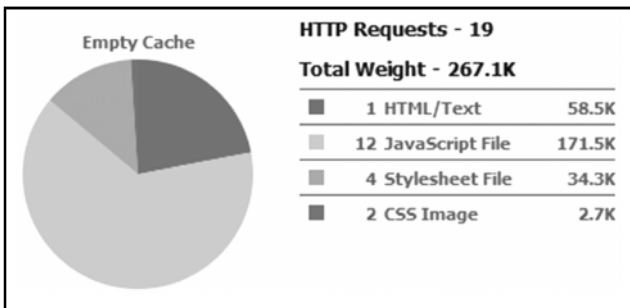


Fig. 2(a) www.usairways.com

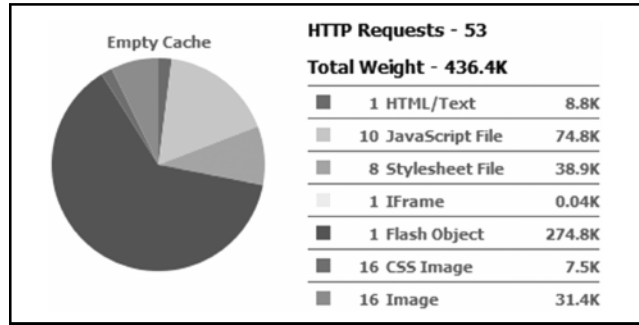


Fig. 2(b) www3.imperial.ac.uk

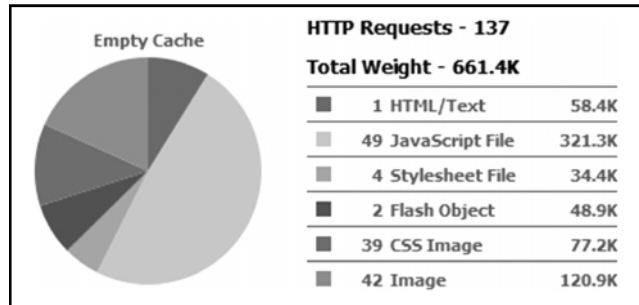


Fig. 3(a) www.usairways.com

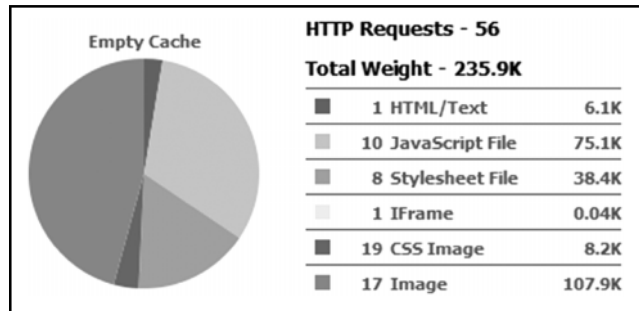


Fig.3(b) www3.imperial.ac.uk

The modifications done to www.usairways.com has increased the number of JavaScript files, CSS images. 2 new flash objects, 42 images are added which further affects the HTTP request number. This is the case when there is tremendous increase in external files. But the pie-chart of www3.imperial.ac.uk shows that even a small increase in external files will lead to rise in number of HTTP requests. We can see that only 3 CSS images and 1 image are added which leads to 3 more HTTP requests.

Modifications to site may be necessary to keep the users keen about the content on your website. Keenness will lead to frequent visits of the users. But, the modifications should not leave the track of optimizing factors. While when we analyzed the corresponding Indian websites again, we found no change in them. This aspect of Indian websites should also be taken care-of. Most of the Indian websites are never modified. Up-to-date websites also have a high impression in user's view which is to be adapted in Indian websites.

#### 4. CONCLUSION

The web traffic has increased so much that now we need to care about the efficient use of the network communication media for storage & retrieval of websites. Web designers and developers should be aware of the optimizing needs and facts for website traffic. Use of proper and to the point data on the website, will definitely lead to improvement in business and efficiency of retrieval of websites. This fact is gradually being understood by those who are in this field for years now. Website development & designing is new in India. This fact of optimization, for improvement in efficiency of website retrieval is still new. Thus, we find that Indian websites lack in many metrics surveyed here. Far more improvement is seen forward in Indian websites as compared to non-Indian websites. Even non-Indian websites have deficiencies which are to be eliminated. In the future, it's a hope that, this paper will at least make the Indian website developers and designers aware that they are lagging.

#### REFERENCES

##### Analyzing Tools

- (a) Yslow – firefox add-ons
- (b) [www.pearanalytics.com](http://www.pearanalytics.com)
- (c) [www.sitereportcard.com](http://www.sitereportcard.com)
- (d) [www.websitegrader.com](http://www.websitegrader.com)
- (e) [www.creatingonline.com/site\\_promotion/website\\_checker.htm](http://www.creatingonline.com/site_promotion/website_checker.htm)
- (f) [www.websiteoptimization.com/services/analyze](http://www.websiteoptimization.com/services/analyze)

##### Paper-content References

- [1] Empirically Validated Web Page Design Metrics- Melody Y. Ivory EECS Department, Rashmi R. Sinha SIMS/ Psychology Department, Marti A. Hearst SIMS UC Berkeley2. Improving Web Site Design- Melody Y. Ivory and Marti A. Hearst University of California, Berkeley.
- [2] Improving Web Site Design- Melody Y. Ivory and Marti A. Hearst University of California, Berkeley
- [3] <http://www.buydomains.com/domain-resources/domain-analytics/page-tagging.jsp>
- [4] <http://www.seoconsult.com.au/off-page-seo/offsite-web-analytics/>
- [5] Optimisation and Accessibility: File Size published by Dan in SEO Basics(<http://www.doublespark-seo.co.uk/blog/optimisation-and-accessibility-file-size/>)
- [6] The International Academy of Arts and Sciences. The webby awards 2000 judging criteria. <http://www.webbyawards.com/judging/criteria.html>, 2000.
- [7] Rashmi Sinha, Melody Y. Ivory, and Marti A. Hearst. Content or graphics? an empirical analysis of criteria for award-winning websites. Submitted for publication, 2001.
- [8] Patrick J. Lynch and Sarah Horton. Web Style Guide:Basic Design Principles for Creating Web Sites. Yale University Press, 1999.
- [9] NomosManagement. WAMMI web usability quuestionnaire. <http://www.nomos.se/wammi>
- [10] Web Criteria. Max, and the objective measurement of web sites. <http://www.webcriteria.com> , 1999.