Study of Indian websites evolution and its importance in web restructuring

Dr. Raj Kulkarni 1, Abhaysinh Sathe2

1,2Walchand Institute Of Technology, Solapur, India

ABSTRACT

Web sites are playing an increasingly important role in today’s society and their economic relevance for the companies is continuously growing. The World Wide Web is in constant evolutionary change in data used to present the web contains and changes in guidelines. These changes must obey the general low of transformation of quantity into quality, which is a general low of any evolutionary process in the world. During the evolution phase the structure of a web application tends unavoidably to degrade. To overcome this degradation we can restructure the web. In this study we investigate the relationship that surrounds these issues, especially those involving the web user interface. This theory of web evolution is not flawless. By understanding these evolution trends we can inform and predict web development into the future. The main purpose of this work is to bring the world trends noticed that new standards and recommendations get adapted faster by the top websites for restructuring.

Keywords: Web evolution, Restructuring, Web crawling

1. INTRODUCTION

Over the last few years we have seen the control in website development shift from programmers to graphic designers. This was a necessary step in the evolution of the industry simply because more companies are seeing the value in maintaining a solid corporate branding throughout all of their marketing pieces. This means that more and more web development is being outsourced by Creative Houses and Advertising Agencies. Web sites are playing an increasingly important role in today's society and their economic relevance for the companies is continuously growing.

The World Wide Web plays vital role in our society – enabling broader dissemination of information and services than was previously available. This World Wide Web is environment of technologies, recommendations and guidelines which are in evolutionary change. These changes must obey the general low of transformation of quantity into quality, which is a general low of any evolutionary process in the world. The transition from web1.0 to web 2.0 is a typically example of this theory.

During the evolution phase the structure of a web application tends unavoidably to degrade. To overcome this degradation we can restructure the web. In this study we investigate the relationship that surrounds these issues, especially those involving the web user interface. This theory of web evolution is not flawless. The growth of the World Wide Web has led to a dramatic increase in accessible information. The web today is a growing universe of interlinked web pages and web apps, teeming with videos, photos, and interactive content. Over time web technologies have evolved to give web developers the ability to create new generations of useful and immersive web. During the evolution phase the structure of a web application tends unavoidably to degrade. To overcome this degradation we can restructure the web. In this study we investigate the relationship that surrounds these issues, especially those involving the web user interface. This theory of web evolution is not flawless.

Commonly two types of method were used to track the changes to the Web. The first method monitors the packet that passes through the corporate firewall in the proxy server [1]. This method can reliably capture the contents from the websites accessed by the server's users, but it is biased towards the needs and culture of the corporation. Thus it does not give a good representation of the Web. The other method (the more popular method) uses a Web crawler or a Web robot to crawl and fetch a snapshot of the targeted WebPages, and the required data for further analysis [9]. The coverage of this method depends on the quality of the web pages. Commonly some kind of webpage selection methodology will be used to select the Web pages like crawling. A Web Crawling was chosen for this study due to the scope, comprehensiveness and flexibility of WebPages it can be programmed to examine, and capture.

To do this, long term slices of a number of popular and randomly selected websites for the last ten years (2002-2012) were captured. This will allow us to investigate if the random websites, in-general, follows the trends of the popular websites, and to identify the lag between them. Two larger sets of popular and randomly selected websites were also captured from the current Web to validate the analysis done for the long term slice of the websites for the last ten years.
This report provides the background material in correlation to the research focus to be undertaken in this study. A detailed discussion relating to the issues of the results were followed. The analysis and discussions covered in this study will contribute as recommendations to inform and predict the Web development into the future.

In this paper we look at web Interaction and restructuring. The paper provides the background material for research focus to be undertaken in this study. The analysis covered in this paper will contribute as recommendation to inform and predict the web restructuring in future.

2. RESTRUCTURING

Maintenance and evolution are critical for website since the requirements often change, the developing cycle is short, while the life cycle is long. In order to make progress in this area, we focus on the users’ responses and attitudes[12]. In accordance to web development procedure the restructuring proposed is differently implemented if it has to do with 1) navigational design (description of the navigational structure of hypermedia application in terms of navigational context like links, indices and guided tours). 2) the change in interface design (i.e. interface changes are built by the web designer like picture, menu, style) [10].

Web restructuring helps to navigate, interact and contribute to the web. For restructuring no homogeneous guidelines are present so that can help to web designer. The guidelines can be used while restructuring web are Authoring tool Accessibility guidelines (ATAG), web Content Accessibility Guidelines (WCAG), User Agent Accessibility Guidelines (UAAG).

Web restructuring focuses on improving technologies that provide interaction with the web. These technologies mainly include (X)HTML which is mark-up language, Cascading Style Sheet (CSS) which provides mechanism for adding presentation style to web page, AJAX a model created to take advantages of the popularity and capability of JavaScript, different types of graphics format like JPEG, GIF, PNG, Flash etc. Development in these technologies effect how people browse the web.

While restructuring use the object oriented technologies so that will be beneficial in future as if web wants the reusability of the contents or structures. For restructuring a website we must know areas for improvement or investigation of a website and we must know what users think of your current site as well as what they want in the website. To understand the need of user and technology adaptation we can use the website evolution analysis. By understanding these evolution trends we can inform and predict web development into the future. The main purpose of this work is to bring the world trends noticed that new standards and recommendations get adapted faster by the top websites for restructuring.

3. METHODOLOGY

Our purpose an understanding of web evolution can be based along two tracks i.e. Web Archives and web Archaeology. Both are important however the archive is far more available as a tool and therefore we will be looking at archival materials. To gather information from web pages we collected a sample of pages from internet archive’s Wayback machine. The Wayback is an archive for web crawls dating back, with snapshot of WebPages organized by date and complete data available. An assumption that a existing website with archives available between year 2001 and 2002 will have archives from at least year 2002 to June 2012. This will reduce the workload of the script and the server to check for a archive in every year. the size and the dynamic nature of the web make it very difficult to select a random sample of web pages. For our analysis we selected a set of 500 Indian websites that covers wide variety of industrial and academic areas as well as geographical region.

4. RESULTS AND DISCUSSION

In this section, we are going to present possible analysis, trends and conclusions from our results. This discussion is described in 8 categories to understand some trends.

HTML Standards:

Decrease in HTML 2 usage for the last ten years was noticed. This was seen for both the random five hundred websites and webby award winning websites. The HTML 3 standards usage also exhibited a decline for the last ten years from 2002 to 2012. HTML 3 is slowly losing its popularity with Web developers/authors. The HTML 4 standards have been heavily used by the Web but a declining trend is predicted. For the XHTML 1.0 standard an increase in usage for this standard was noticed. Thus for the top websites, it was observed that a major shift in usage (> 50%) from HTML 4 to XHTML 1.0 had already occurred. This converting trend is expected to continue as it can be verified by the discussion presented earlier for the individual HTML standards trends.
On average the Webby award winning top websites adopts to a new standard one year faster than the random websites, and a growth in XHTML 1.0, and 1.1 usage was predicted. A trend was also notice that the usage of HTML 4 is decreasing, and websites that were previously using it are replacing their WebPages with XHTML 1.0.

As shown in the below graph the evolution of HTML 4 usage percentage from 2002 to 2012 for Webby as well as random 500 websites. The usages of HTML 4 in webby award winning websites and random 500 websites are nearly same except the usage between 2006 to 2008 which is more in webby award winning websites.

Figure 1: HTML 4 usage percentage for Webby Award Winning Websites and Random 500 websites

Image Technology and Graphics

There are many different types of graphical formats available to be used over the Web, however as mentioned earlier, only JPEG, GIF, SVG, PNG, SMIL and Flash formats will be covered. Using the Webby top websites and the random five thousand websites results for verification, the usage of GIF was expected to remain unchanged for the near future. Webby top websites and the random five thousand websites results for verification, the usage of GIF is constantly increasing. The growth of JPEG usage in Webby websites and random 500 websites nearly similar. From the results of the Webby top websites and the random five hundred websites, it shows a growing trend for the usage of PNG.

The use of streaming media on the Web has increased by more than 100% each year (Li et al. 2005) [7]. From 2002 to 2005 the total volume of streaming media files stored on the Web grew by more than 600%. More than 87% of all streaming media is abandoned by users in the first 10 seconds, however, wasting up to 20% of server bandwidth. While only 3% of server responses are for videos, they account for over 98.6% of the bytes transferred [8]. In true Pareto fashion, about 10% of the most popular videos on YouTube account for nearly 80% of the views, making caching an appealing performance enhancement.

Figure 2: GIFs usage percentage for Webby Award Winning websites and top 500 websites

Based on our current Web analysis for the Webby top websites and the random five hundred websites, this trend is predicted to continue. On the average it will take about two years for a new graphical format to get adopted by the Web from the time it was released. As shown in figure 2 the usage of GIF images in webby award winning website is generally more than random 500 websites. But the decrease in usage of GIF images is predicted.

In last five years the usage of JPEG images is more in random 500 websites but showing decreasing trends as shown in figure 3.

Figure 3: JPEG’s usage percentage for Webby Award Winning websites and top 500 websites
Cascading Style Sheets:

A steady growth was noticed and we predict that this growth will continue for the next year. When analysing the trend for the usage of CSS to be more than 50%, although both showed similar trends, a four years lag was noticed between the top websites and the random websites. Hence for this type recommendations to get adopted by more than 50% of the Web, an adoption time of about four to nine years is required, and less than four years for Web technologies that surrounds it to get developed. CSS allows the Web developers/authors the flexibility, and more control over the webpage's presentation, thus this led to an increase in the usage for different types of graphical formats such as JPEG.

These trends suggest that the take up of JPEG may have benefited from the increase of the usage of CSS. CSS has given the web the flexibility and the control over the presentation of web contents. Because of this a healthy usage growth for this standard was seen and graphical formats such as JPEG usage also benefiting from it.

![CSS usage percentage for Award winning websites and 500 websites](image)

AJAX:

AJAX is a model created to take advantage of the popularity and capability of JavaScript, the asynchronous technology, and XML [5]. It can be noticed that even when the usage of AJAX was increasing, a decline in JavaScript usage was still observed. This showed that the AJAX model may be gaining popularity, but it may not be enjoying the full increase, but it was sharing it with other technologies such as Flash that is capable of providing the asynchronous model. As seen in figure 6 usage of AJAX technology was shown increasing trend till 2008 but after that the ratio of inclined is decreased.

![AJAX usage percentage for Webby Award Winning Websites and Random 500 Websites](image)

E JavaScript:

It was observed that most of the Webby award winning top twenty websites uses some JavaScript in their website design, but a decline in usage was also noticed since July 2007. Our Random five hundred websites results also demonstrated similarly trend. By using our Webby award winning websites and the random five hundred websites results to validate these claims, a continuous decline in JavaScript usage was predicted. Further analysis is required to understand more about what causes this decline. As the usage of Flash is increasing, it is resulting in declining in usage of JavaScript. At the same time AJAX getting popularity due to cooperation of Flash. It is seen that the usage of JavaScript war nearly 100 percent in the Webby award winning top 20 websites in year 2006. As seen in figure 7 the usage of JavaScript is showing continuous declined trend from 2005 in case of Webby award winning websites.

![E JavaScript usage percentage for Webby Award Winning Websites and Random 500 Websites](image)
Figure 6: JavaScript usage percentage for Webby Award Winning Websites and Random 500 Websites

F  Web Server:

As one would expect the Open Source Apache Server is by far the most popular server. It has held this position since April 1996 according to Wikipedia. While this is an indication of the Apache version distribution, the majority of Apache sites do not advertise the exact version number. Apache version 1.3 is deprecated and no longer maintained. Now Apache 2.2 is popular web server. Apache 1.3 is more popular than apache/2. IIS is the second largest popular web server. IIS 6.0 is clearly the most popular. if there are sites still running IIS 3.0 or 4.0 these sites will be running on Windows NT4.0. IIS 5.0 will be running on Windows 2000 servers. Generally IIS 6.0 was the first version released by Microsoft where Security started to become a priority. IIS 8.0 and 9.0 have not been released so are also likely user edited server strings..

G  Flash:

It was noticed that around 2007, an increase in Flash usage was noticed around the same time when JavaScript began to roll off. This analysis supplies a reason for the trend of the increase in AJAX popularity, and the contrary results for the usage of JavaScript trend. This showed that the AJAX model may be gaining popularity, but it may not be enjoying the full increase, but it was sharing it with other technologies such as Flash that is capable of providing the asynchronous model. It can be noticed that even when the usage of AJAX was increasing, a decline in JavaScript usage was still observed. Therefore another analysis was carried out between Flash and JavaScript as seen in figure 9. The rate of increase in Flash usage for webby top 20 websites is more than that of random 500 websites.

Figure 7: Flash usage percentage for Webby Award Winning websites and top 500 websites

H  PHP:

In last 10 years PHP 3.0, 4.0 and 5.0 are used for web development. After the release of PHP 3, usage really started to t changes to the architecture of PHP. These changes included abstracting the layer between the language and the web server, adding a thread-safety mechanism, and adding a more advanced, two-stage parse/execute tag-parsing system. Currently PHP version 5.2 is used which is the most popular PHP version among other PHP version.

Evolution from Web 1.0 to 2.0:

Most website were using HTML and static pages from 2003 onwards web 2.0 using XML and AJAX for web development. The communication architecture in web 1.0 was one to one in web 2.0 it was shifted to many to many. Structure of website in web 1.0 was decorative pages that are made to view through web browser. In web 2.0 web structure is like many shared content to be shared with interactive services. Visualization was browser in web1.0 in 2.0.
5. Conclusion

Web page design and layout techniques have undergone many changes over the last 10 years. In this study we look under the hood of pages and examine the evolution of Webpage usage from 2002 to 2012. The discussion on the analysis presented above highlights and predicts possible trends for individual standards, and recommendations. In this paper we need to understand the evolution of infrastructure, guidelines, and recommendations that are useful for web restructuring.

By understanding the way the Web has evolved we can attempt to predict how it will evolve, and as such guide our future efforts in supporting Web Restructuring. From this study we will used to understand the relationship and trends between the underlying web Standards, recommendations, guidelines, and its adoption time which will help in web restructuring.

References


[12] Lei Xu; Baowen Xu; Zhenqiang Chen; Huowang Chen;Dept. of Comput. Sci. & Eng., Southeast Univ., Nanjing, China “Website evolution based on statistic data” in Distributed computing system 2003,FTDSC.proceedings yhe ninth IEEE worpkshop on future trends of distributed computing system.

