

Categorize Web Sites Based on Design Issues

Amin Rasooli^{1,2}, Fattaneh Taghiyareh¹, and Peter Forbrig²

¹ School of Electrical and Computer engineering, University of Tehran, Tehran, Iran

² Department of Computer Science, University of Rostock, 18051 Rostock, Germany

amin.rasooli@uni-rostock.de, ftaghiyar@ut.ac.ir,
peter.forbrig@uni-rostock.de

Abstract. Interface design is one of the most important topics during web development process. The final design is a tradeoff between the owner's personal idea and the web developer's perception of what he wants. In this paper, we have proposed a new model called WLDM (Web layout design model) to cover the important components of interface design. There are three components in the WLDM, including structure, content and visual. We have selected three features for structure, two for content, and three for visual component. Thereafter, we have made a dataset using 1088 most visited web sites. Finally, applying K-means algorithm, we have clustered this dataset. According to our result, six clusters were identified. Considering WLDM, web layout designer have a blueprint to cover areas of research related to this issue. The result of this clustering can be used for recommender systems to map owner groups, which have different attitude.

Keywords: WLDM, Web design, Adaptation, Personalization, Layout, User Interface, Human-Computer Interaction, Clustering, Data mining.

1 Introduction

Web design is the task of making presentation of content, which are delivered to end users through the Internet network by using a Web browser or other Web based software. "About 40% to 70% of a web site's code is its user interface".[1] The goal of web design is to make a collection of some documents and elements that put on a web server or more, and illustrate content and user interfaces to the visitors in form of some hypertexts. Some items like text, images, and forms can be put on the page using HTML tags. Displaying more complex media like graphics graphs, animations, videos, and sounds needs some plug-ins such as Adobe Flash, Quick Time and so on. [2], [3]

Despite the varieties and volumes of web site templates [4], the owners are looking for a more principled and attractive layout for their web site rather than other web sites. In this situation, each part of webpage design can have an essential role in the success of a web site, but this process is not always efficient and adjustable for maximize usability of the web site; it may be inefficient and will be totally different to his taste.

Among the reasons for aforesaid condition, we can mention the incompatibility between the web site owner's needs and the final implemented web site. In other words, the problem is lack of the web site manufacturer's understanding of owner's needs and his visitors' needs the conditions described by owner. This phenomenon leads to making a product, which do not satisfy anyone. This kind of problem is usually happens to the owner who does not have enough information about the web design, or sometimes he cannot present what he exactly wants (actually it results from the lack of the owner's knowledge about web site or the lack of designer's familiarity with situation). Therefore, if an owner knows exactly what he wants, then designing companies also understand his expectation, then they can help him to have a suitable web site. Otherwise, they will encounter many problems either in the initial design or in delivery of the final product. [5]

We started by introduction of our aims then present model of web interface features in Section 2. In Section 3, we did clustering with K-mean algorithm on 1088 web sites. Finally, we will summarize contributions with suggestions for future research in Section 4.

2 WLDM Model

Our proposed model is for supporting and adapting design personalization that we called it as WLDM (Web layout design model). In the following, we can see the usage of this model in order to improve design of the web site's layout. [6]

Different components of the web site are mentioned, as you can see in Fig. 1. It can be divided into three different categories: structure, content, and visual. We will discuss in details about these three components and their branches, in follows. [7][8]

2.1 Structure

Frame and formation of a web site is related to structure components. The structure of web site made up from different parts, we consider these items: types of web interface, links and cognitive load.[9]

Types of web interface. Web sites can be categorized in different types. Each one is specialized in a particular service or usage. Usability, visualization, and functionality are three distinct issues, meanwhile they related on type of web site. Six major types according to some researches which, are Graphical, Menu driven, Search engine type, Discussion board type, Content managed web site and Flash web site.[10]

Links. We analyses links from two standpoints, on one hand place and appearance of links, And on the other hand navigation and target issues.

Replacing links in a prominent place for example on the main body of web site and putting similar items close to each other. Only have unique navigation system, not different way for specific path. Using icons is an option for representing links in web site but it is better to considering relation between links and the images. [11]

Distinction of links in a web site is the idea that based on some links lead to general information and some of them lead to detailed information. It should be considered that if it is better to put all the content (for example for news) or only a link for head line. [12]

Finding the intended item for the left presentation position was appreciably faster than that of the bottom and top positions. In addition, for the right position found to be notably faster than the bottom position. Menu items should be placed vertically in order to decrease speed of finding a word. [13] And horizontal menus are searched more quickly than vertical ones. [14][15][16]

Cognitive load. A common problem for computer users is Learning and remembering the information which is presented on a computer screen. Based on cognitive theories, one of the reasons for lacking the retention is user's inability to create a mental schema, picture, or plan about the information is illustrated on a computer screen. To create a schema or plan, users should find out, which an obtained knowledge is related on a big picture.

If you are lost on a web site several times, there must be a problem and confusing informing a schema so it should be considered seriously for better retention by users. We can conclude that more information cannot cause more cognition, the ways of providing the information can make a schema, shape, or plan for better understanding better. [17]

2.1 Content

In general, “content” contains everything in your web. This may involve documents, data, applications, service. Moreover, if we had determined that a web site had not completed by the qualified stuff, we should logically progress to evaluating the properness of web sites. [9] Four issues of content investigation is as follows. [18]

Link Suggestion. “Reduce the number of links to ensure that genuine and necessary links are clearly identified”. [12]

Cognitive load and impairment of learning can be the result of higher number of links for hypertexts per page. Several links as a navigation support techniques are recommended for leading learners and reducing cognitive overload practically. It is recommended to use hypertext design for accessing the information in an easy way and can help to improve user understanding on cognitive issues and so on can cause increasing in learning.[19]

Hypertext design is directed for enabling information access in a simple way and leading to an improvement of readers' understanding of cognitive stuff and increasing in learning.

Content faults. When we redesign or change the web site, it can cause some faults which should be considered and eliminated. Finding faults and fixing them can be classified according to type of service, failure identifiers, fault analyzers, analysis and repair tools. For more information we can discuss about the some of them as following;

- Failure identifiers (some failures could be discovered by user actions simulation, for example filling a form).
- Fault analyzers (it found out the failures and highlighted them).
- Analysis and repair tools (they help fixing the faults). [20][21]

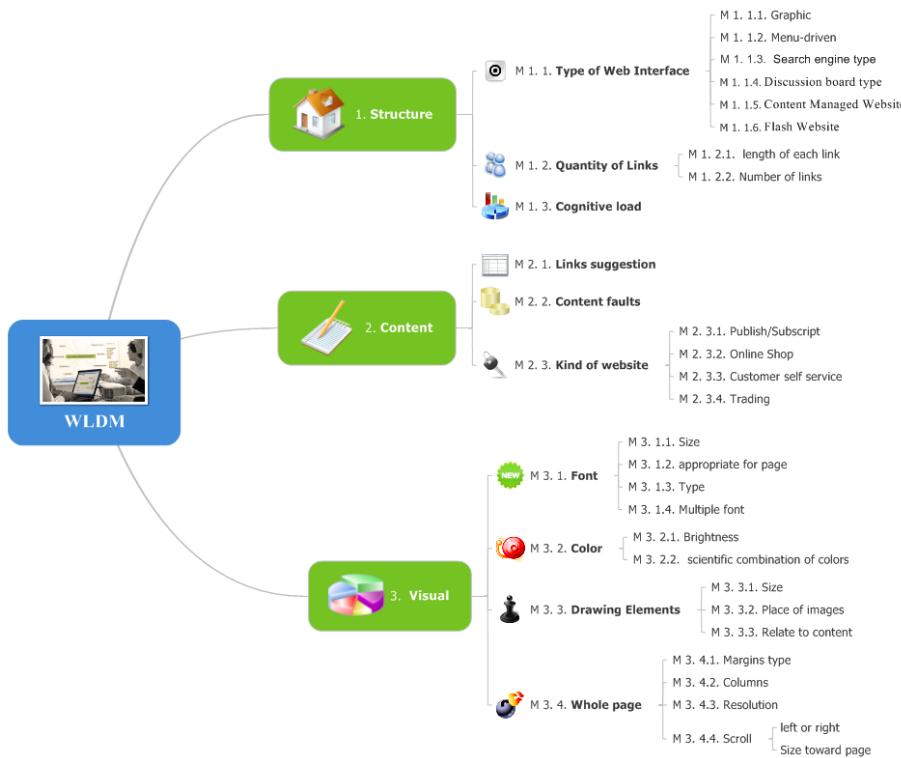


Fig. 1. Personalization Model of important components in a web site

Kinds of web site. In our study, we used the web site's type, which adopted the compact IBM classification of Web sites according to volume of traffic. Based on criteria such as: pages retrieved, number of transactions, their complexity, type, and number of searches, information stability, and security concerns, this classification proposed five types of high-volume Web sites: publish/subscribe, online shopping, customer self-service, trading, and B2B. [1]

2.3 Visual

Visual items aim to suggest a basic concept and prototype. If users want to experience personalized page design, they should express their preference about page design elements such as font, color, drawing elements and the whole page. Then, the personalization system simply assembles design elements selected by the user into a webpage and displays the result. We consider the element in detail.

Font. A significant element in designing webpage is the font and its attributes like size, type and so on. These attributes are very important because of their affect on the speed of reading and could attractive users. Some significant differences in reading efficiency were detected between the font types.

In general, Times and Arial type are read faster than Courier, Schoolbook, and Georgia. Fonts at the 12-point size read faster than fonts at smaller or bigger size. For font attractiveness, Georgia is recognized as being more appealing than Arial, Courier, and Comic, while Times is recognized as more appealing than Courier. [22]

Generally, older adults are very sensitive to font size on a page. The senior adults were changed with performing tasks such as finding special information on a page. [23]

Color. Color is one of the most worthy items in psychological aspect in this research area. There is a close relationship between colors of web sites and emotions. Web site colors can effect on Users' behaviors and cognitive processes especially when they are associated with a specific physical layout. A few recommendations are useful for appropriate color selection for designing user interfaces. It could be comprises of some aspects of interfaces like text background. In visual aspect, harmony is related on a suitable arrangement of different parts and some while it is enjoying by the eyes. In a web site, people should feel a sense of equilibrium and pleasure.

Color has different meaning in different culture and different nation have different expectation, and belief based of the meaning of each colors. So each color has its own psychological influences in any web site.

Drawing elements. Drawing elements can affect the web site in different aspects. Using pictures with high resolution increase the size of the web sites. Then for each picture, appropriate format and size is needed. Pictures and icons should be in harmony with the web site's type, background and other details of the web site.

Meanwhile picture's brightness, place of pictures and being related to the content are the important criteria should be considered.

Whole page. Total view, web site's margin, diagrams, and resolution are some of the parameters in this area. Diagrams and columns should be symmetrical otherwise, it is important to follow specific rules. The web site's resolution depends on the user's browsers and should get regulated with them.

Another point that should be taken into the consideration is scroll. Moreover, it is required not to use vertical scroll in the web site. It is often cited as evidence that increasing line spacing decreases reading time. This finding is attributed to the fact that close spacing of text requires more "eye fixations per line" and, as a result, fewer words are read during each eye fixation. [24]

3 Data Gathering from Web Sites and Clustering

One of the aims of this article is to cluster the web sites based on the mentioned features, which is selected by considering the eight important issues from all of them. These features are:

- Cognitive load: It would measure by sum of the lines and words in the homepage.
- Type of web interface: It consists of six types but we consider three of them such as normal html, graphical (or flash) type, and the other types. These types investigate by detecting words in introduction of the web site.

- Complexity: It related to the time of decoding and compiling the web site in some parts by our software. “a” consider as the low complexity, “b” as middle, and “c” as the high complexity.
- The quantity of the Links: Number of the links in the homepage.
- The quantity of the Images: Number of the images in the homepage.
- Web site’s goal: It has got four options: Publish/Subscript, Online Shopping, Customer self-service, and Trading. They have been recognized with analogy issues, selecting, and mapping significant words in the web site.
- Font variety: font changing quantity.
- Color variety: color changing quantity especially in text.

In Table. 1, We can see the parts of the input of the program.

Table 1. Parts of Automatic Production of Web sites’ Attributes

Brand	Cognitive load	Complexity	Type of WI	links	images	Web site’s goal	Font	Color
google.com	741	a	HTML	17	6	Customer	11	47
live.com	3054	b	HTML	51	16	Publish	3	7
yahoo.com	7319	b	etc.	53	23	Online	14	162
microsoft.com	3596	b	HTML	71	15	Customer	2	0
facebook.com	1498	a	HTML	14	5	Customer	0	0
youtube.com	1201	b	Graphical	39	75	Publish	0	2
wikipedia.com	2768	a	HTML	33	51	Publish	10	3
aol.com	3463	c	etc.	49	45	Online	3	7
ebay.com	1757	b	Graphical	89	91	Online	-1	5

The software, which is used for extracting these data, has been written in PHP [25]. We investigated more than thousand of the popular web sites [26][27]. Next step was cleaning and preprocessing the data, and after preparing data, we used a popular data mining software, which is called Weka 3.6.3. The Weka workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality [28]. We used the K-means method. Consequently, web sites are divided into 6 clusters [29].

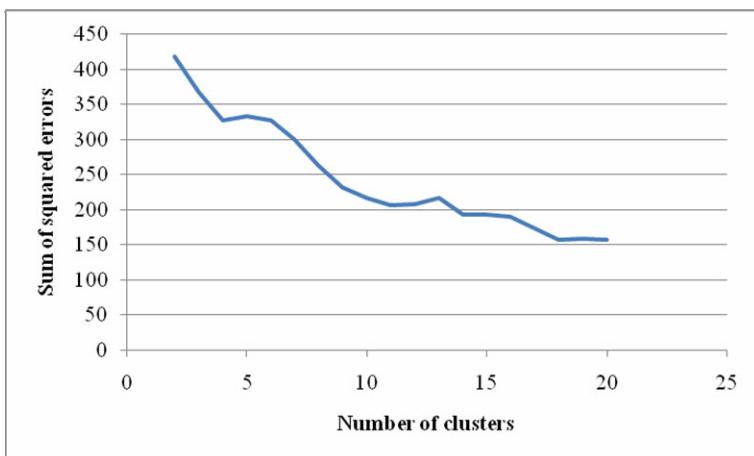
The result of Web site clustering is exhibited in Table .2. These web sites were clustered according to the above-mentioned features. We had considered K between 2 till 20 as the quantity of the clusters by k-means algorithm. For four of the clusters we got optimum result.

The result of the clustering is illustrated in Table 2. Missing values were replaced with mean/mode. Within the clusters, the sum of squared errors is 367.83504. In addition, Fig 2 illustrate the vector of sum of squared errors by different amounts for k. Regarding to this graph (Fig. 1) we can see the largest drop in the sum of the within sum of squares in k=4 as the best quantity number of our clusters.

Table 2. Clusters centroids

Attribute	Full Data (1088)	0	1	2	3
Cognitive load	63423.5441	110945.378	41849.9167	36468.5273	60300.8519
Type of WI	html	graphical	Html	graphical	html
Complexity	simple	complicated	simple	Complicated	complicated
Quantity of links	28.5625	56.54756	19.94352	6.74	22.5
Quantity of images	C	c	a	b	a
Web site goal	Trading	Online	Online	Online	Customer
Font variety	1/4	1/4	1/4	1/4	7/10
Color variety	21.4853	35.1585	14.5278	9.5818	32.037

This part of this study can be used as follows; first of all the owner will be classified, then each group will be mapped to our clusters as a class column, which is mentioned in Table 2 [30]. Finally, a new model is created. When new owner wants a layout, we can map him to one of the owner's group. The methodology and results help identify new layout by using collaborative filtering in recommender systems for the new owner [31][32].

**Fig. 2.** Sum of squared errors by different amounts for k (between 2 until 20)

4 Future Works and Conclusion

This study comprises of two stages, the first stage was collecting the information about the important issues in web designing, and the result has been presented as a model named WLDM.

In the second stage, we select eight important items from WLDM's model, which could be measured, and 1088 web sites have been analyzed by considering them and the results of them have been cauterized.

The outcome of our research can be useful for designing the desired web sites, which saves time and price during the design process, our results can be used to propose an efficient and effective web designing to attract more visitors. From the subcategories of three primary components in the Fig. 1 we have used eight features which are more measureable, such as: Cognitive load, Type of web interface, Complexity, The quantity of the links, The quantity of the Images, Web site goal, Font variety and Color variety. The methodology and results help identify user interface issues that a designer should emphasize. Also from the second part of this study, we considered which one of the mentioned results have been used for designing by the world famous web sites..

We provided these data in order to use in third part, which we did not write in this paper and the basic mechanism behind collaborative filtering systems for future work of this research is in the following: A large group of owners' preferences are registered. Using a similarity metric, a subgroup of owners is selected whose preferences are similar to the preferences of the person who has same attitude. Average of the preferences for that subgroup is calculated. We can build a model by data classification and could map the owner group to web sites clusters. In addition, the final software recommends the layout design to new users. In our future work, the model will be used for recommender systems that allow to replace web site designers.

The system with scrutiny of owner's behavior and with the use of psychological methods, suggest him dedicated desired patterns. One of the methods that can be used in this system is investigating the owner's behavior and compares it with another owner using these components. The final aim is making a recommender system as an automatic web designer.

Acknowledgments

We thank the Iran Telecommunication Research Center for financial support.

References

1. Zviran, M., Glezer, C., Avni, I.: User Satisfaction from Commercial Web Sites: The Effect of Design and Use. *Information and Management* 43(2), 157–178 (2005)
2. Wikipedia (October 2010), http://en.wikipedia.org/wiki/Web_design
3. Parons, J.J., Oja, D.: *Computer Concepts*, Course Technology, Cambridge, MA (1998)
4. Kazienko, P., Adamski, M.: AdROSA—Adaptive personalization of web advertising. *Information Sciences Wrocław*, Poland 177(11), 2269–2295 (2007)
5. Garrigós, I., Meliá, S., Casteleyn, S.: Personalizing the Interface in Rich Internet Applications. In: Vossen, G., Long, D.D.E., Yu, J.X. (eds.) WISE 2009. LNCS, vol. 5802, pp. 365–378. Springer, Heidelberg (2009)
6. Seo, J., Lee, K.: Development of Web site Design Personalization Service Using Design Recommender System (2003), <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.108.3769> (access time: October 2009)
7. Shapira, B., Taieb-Maimon, M., et al.: Study of the usefulness of known and new implicit indicators and their optimal combination for accurate inference of users interests. In: ACM Conference (2006)

8. Heimlich, J.E., Wang, K.: Evaluating the structure of web sites. EETAP Resources Library Publications 2 (January 2002)
9. Srivihok, A., Sukonmanee, P.: Intelligent Agent for e-Tourism: Personalization Travel. In: Proc. of The 14 th International World Wide Web Conference Workshop (WWW), pp. 10–14. Keio University, Chiba (2005)
10. Nielsen, J., Tahir, M.: First page Usability 50 Web sites Deconstructed, New Riders, Indianapolis (2002) ISBN
11. Evett, Brown, Evett, L., Brown, D.J.: Text formats and web design for visually impaired and dyslexic readers: clear text for all. *The Journal of Interacting with Computers* 17, 453–472 (2005)
12. Tullis, T.S.: Screen design. In: Helander, M., Landauer, T.K., Prabhu, P. (eds.) *Handbook of Human–Computer Interaction*, pp. 503–531. Elsevier Science, Oxford (1997)
13. Scott, D., Findlay, J.M.: The shape of VDU's to come: a visual search study. In: Lovesey, E.J. (ed.) *Contemporary Ergonomics*, pp. 353–358. Taylor & Francis, London (1990)
14. Foltz, P.W., Kintsch, W., Landauer, T.K.: The measurement of textualcoherence with Latent Semantic Analysis. *Discourse Processes* 25, 285–307 (1998)
15. Norman, K.L.: Better design of menu selection systems through cognitive psychology and human factors. *Human Factors* (Golden Anniversary Special Issue) 50, 556–559 (2008)
16. Tracy, J.P., Albers, M.J.: Measuring Cognitive Load to Test the Usability of Web Sites. In: Proceedings of the Annual Conference for Technical Communication, pp. 256–260 (2006)
17. Schiaffino, S., Amaldi, A.: User–interface agent interaction: personalization issues. *International Journal of Human–Computer Studies* 60(1), 129–148 (2004)
18. Ignacio Madrid, R., Van Oostendorp, H., Puerta Melguizo, M.C.: The effects of the number of links and navigation support on cognitive load and learning with hypertext: The mediating role of reading order. *Computers in Human Behavior* 25(1), 66–75 (2009)
19. Nygren, E.: From paper to computer screen: Human information-processing and user interface (1996)
20. Chalmers, P.: The role of cognitive theory in human-computer interface. *Computers in Human Behavior* 19, 593–607 (2003)
21. Ling, J., van Schaik, P.: The influence of font type and line length on visual search and information retrieval in web pages. *Int. J. Hum.-Comput. Stud.* 64(5), 395–404 (2006)
22. Micheal, B., Lida, B., Riley, S., Hackler, T., Janzen, K.: A Comparison of Popular Online Fonts: Which Size and Type is Best? *Usability News* 4(1) (January 2002)
23. Weller, D.: The Effects of Contrast and Density on Visual Web Search, <http://psychology.wichita.edu/newsurl/usabilitynews/62/density.asp>
24. Welling, L., Thomson, L.: PHP and MySQL Web Development, 3rd edn. Sams Publishing, USA (2005)
25. SuperPower: Visualising the internet (January 2010),
http://news.bbc.co.uk/nol/shared/spl/hi/sci_nat/10/internet_treemap/assets/nielsen_top_100_january_2010b.xls
26. The 1000 most-visited sites on the web (December 2010),
<http://www.google.com/adplanner/static/top1000/>
27. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. *ACM SIGKDD Explorations Newsletter* 11(1), 10–18 (2009)
28. Aittokallio, T., Malminen, J., Pahikkala, T., Polo, O., Nevalainen, O.: Inspiratory flow shape clustering: An automated method to monitor upper airway performance during sleep. *Comput. Methods Progr. Biomed.* 85, 8 (2007)

29. Cho, Y., Kim, J., et al.: A personalized recommender system based on web usage mining and decision tree induction. *Expert Systems with Applications*, World Scientific and Engineering Academy and Society (WSEAS) 23(3), 329–342 (2002)
30. Melville, P., Mooney, R., et al.: Content-boosted collaborative filtering for improved recommendations. AAAI Press; MIT Press, Menlo Park, CA (2002)
31. Huang, K.J., et al.: A survey of E-commerce recommender systems. *Service Systems and Service Management*. In: International Conference on In Service Systems and Service Management, pp. 1–5 (2007)
32. Shapira, B., Taieb-Maimon, M., et al.: Study of the usefulness of known and new implicit indicators and their optimal combination for accurate inference of users interests. In: ACM Conference (2006)
33. Stormer, H.: Personalized web sites for mobile devices using dynamic cascading style sheets. *International Journal of Web Information Systems* 1(2), 83–88 (2005)
34. Card, S.K., Moran, T.P., Newell, A.: *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum, Hillsdale (1983)
35. Sheppard, S., Rouff, C.: In: John Marciniak, J. (ed.) *Encyclopedia of Software Engineering*, vol. 2, John Wiley & Sons, New York (1994)
36. Kintsch, W.: The role of knowledge in discourse comprehension: A construction integration model. *Psychological Review* 95, 163–182 (1988)
37. Templatemonster (October 2010), <http://www.templatemonster.com>
38. Nahm, E.S., Preece, J., Resnick, B., Mills, M.E.: Usability of health web sites for older adults: A preliminary study. *CIN: Computers, Informatics* 22, 326–334 (2004)
39. Kolers, P.A., Duchnicky, R.L., Ferguson, D.C.: Eye movement measurement of readability of CRT displays. *Human Factors* 23, 517–527 (1981)
40. Bernard, M., Liao, C.H., Mills, M.: The effects of font type and size on the legibility and reading time of online text by older adults. In: CHI 2001 Extended Abstracts on Human Factors in Computing Systems, CHI 2001, Seattle, Washington, March 31 - April 05, pp. 175–176. ACM Press, New York (2001)