J.T. Yao, V.V. Raghvan, G.Y. Wang (Eds.): WSS'04, pp. 94-101, 2004.

The Design of an Integrated Web Mining and Usability System

Ringo Wai-kit Lam, Chun-hung Li, Chun-kit Chui
Department of Computer Science
Hong Kong Baptist University
chli@comp.hkbu.edu.hk

Abstract

Web usability is an important and sometimes controversial research area. We review the different approaches to web usability and illustrate that the factors influencing Web usability are often incompletely analyzed. We proposed an integrated system for web mining and usability study where four core modules are designed to address the fundamental issues in usability analysis. The integrated approach allows a totalistic view of the web usability and facilitates analysis across different modules. As an example to cross modules analysis, we apply association rule mining from the link structure obtained from web mining module to automatically discover menus and structures in a web site. Furthermore, such mining tools allow the decoupling of the design-based link structure from the contextual-based link structure.

1. Introduction

Usability test is always regarded as a costly and time exhaustive process but it is becoming prevalent and more important as many of our business operations and social activities are processed and completed on the Internet directly.

Usability evaluation is usually conducted in a number of ways, e.g. user testing, heuristics evaluation and automatic tool analysis. Since usability evaluation is very expensive, automatic tools were developed in the last few years to help the web designer evaluate the web site. However, such automatic tools still cannot replace the value of testing with the actual users.

An integrated system that enables Web Mining and Usability analysis (**Webmius**) is proposed in this paper. The system is composed of four modules that perform the function of the task-based usability evaluation, the web design and layout analysis, the web structure mining and the semantics inference.

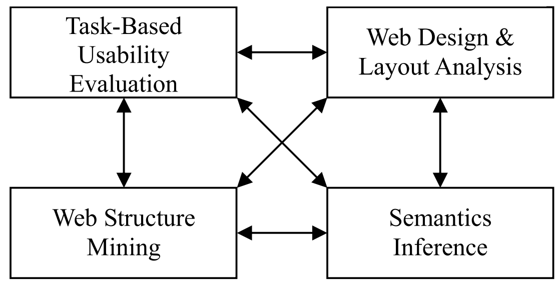


Fig.1 Web Mining and Usability System

Conventionally, web usability analysis is focused on the function of either one or two of the modules we implement. For example, for task-based usability evaluations, studies include [26, 27]. Design and layout analysis includes [17]. Web structure mining has [2, 21]. For semantics inference, there are [6, 9, 10].

The four modules shown in Fig.1 are integrated to tackle the usability evaluation. Except the task-based usability evaluation module, the other three modules may run automatically. The task-based usability evaluation module requires actual user involvement although simulated users like Bloodhound [7] and MESA [24] have been developed, the simulated users are still not mature. On the other hand, metrics are being developed for the other 3 modules. These metrics try to capture the web design parameters, such as color, font, layout, image size, etc. Some of the metrics are similar to [17] but our metrics are also emphasized on the hyperlink and menu structure as well as their positions.

Concerning the design and the layout, research study [14] found that the search time of a hierarchical and labeled layout was faster. Besides, McCarthy et al. [23] reported that users rapidly adapted to an unexpected screen layout and the internal consistency of a web site was the most important. Ken Hinckley [12] extended Fitts' Law to consider the IBM ScrollPoint and the IntelliMouse Wheel. Their experimental approach revealed a crossover effect in performance versus distance, with the Wheel performing best at short distances but the ScrollPoint performing best at long distances.

This paper first describes the background of usability metrics and testing tools. Then the characteristics and the design of our system are discussed. Our Webmius platform is described in section 5. Finally, an approach using association rule mining to remove the menu is introduced to simplify the web site structure for further analysis.

2. Usability Metrics and Evaluation Tools

There are many metrics and evaluation tools. Bobby, A-Prompt, WebSat, 508 Accessibility Suite and WebTango were developed to enable designers and evaluators to verify a page or a web site according to the guidelines [4]. Their main problem is that the guidelines are hard coded in the tool.

Ivory summarized the page characteristic metrics of the other studies in [15] that generally classified the metrics into Page Composition, Page Formatting and Overall Page. Besides, the EvalWeb project http://lis.univ-tlse1.fr/evalweb/ tried to create a framework to organize the guidelines to help people structure the guidelines and use them for design and evaluation of web sites. Guidelines were classified into 5 categories : (1.) Design rules, (2.) Ergonomic algorithms, (3.) Style guides, (4.) Compilations of guidelines and (5.) Standards.

Usability.gov that is maintained by the U.S. Department of Health and Human Services (HHS) also put over 50 Web design and usability guidelines on their web site. IEEE has a Std 2001-1999 [http://www.computer.org/cspress/CATALOG/st01117.htm] which defines recommended practices for web page design and implementation for intranet/extranet environments. The American's National Institute of Standards and Technology (NIST) [25] on the other hand developed a prototype tool WebSAT for researching usability rules. It allowed either to use its own set of usability rules or those of the IEEE Std 2001–1999.

Apart from the metrics or guidelines for the usability test, there are over 50 evaluation tools developed [16]. One of the major tools is WebTango [17]. Under the WebTango project, the 157 measures were categorized into 9 major types, (1.) text elements, (2.) link elements, (3.) graphic elements, (4.) text formatting, (5.) link formatting, (6.) graphic formatting, (7.) page formatting, (8.) page performance and (9.) site architecture. The assessment was separated into site-level and page-level. However, the site-level assessment was not comprehensive enough, it only reflected the total number of pages, the breadth and the depth traversed by the crawler.

To mine the user behavior from the task-based usability evaluation, logging tools are commonly implemented. A recent project which embeds logging design is TEA [26]. It is an open source project that develops a client-side proxy to capture the client-side events and feedback to the analysis server. WebRemUSINE [27] is similar to TEA but no client-side proxy is set up. WebRemUSINE's logging tool is able to capture the browser's event by event handler scripts. Unfortunately, the scripts are not persistent. Each page of the site has to include the script and all events are communicated to the applet that sends the server back with all the logged events at the end of the session.

WebQuilt [13] was built by the User Interface Research Group at the University of California at Berkeley. A tailor-made Java-based proxy server was developed to record the link structure of the web pages. However, the WebQuilt has its shortcomings, it cannot handle Flash, Java Applet or any kind of web page which has links or redirects created dynamically by JavaScript and other browser scripting languages cannot be handled. As a consequence, the JavaScript generated pop-up windows and DHTML menus popular on many web sites are not captured by the WebQuilt.

Heer et. al. [11] introduced an evaluation by building user profiles and combining users' navigation paths with features, such as page viewing time, hyperlink structure, and page content. They tried to find how well these features contributed to the clustering process in real world and to evaluate whether the clustering algorithms correctly categorized the user sessions so that the real user's behavior might be determined from the web log.

They used WebQuilt proxy-based logger [13] to capture all of the user sessions, therefore their system might suffer from the drawback of WebQuilt's deficiency. On the contrary, our work extends the work done by [11] to include a more comprehensive set of features and system design. Besides, we also consider the hyperlink and menu structure, and their positions in our system.

3. Web Site Structure

Conventionally, a web site structure may be evaluated with the Card Sorting technique; however, this technique is difficult to implement for a large and information-centric web site. There are a number of commercial and free evaluation tools available. Most of the tools are based on the user logging results that are stored in the proxy server or through the embedment of client-side programming code to capture the client's events.

A recent publication by Miller and Remington [24] pointed out that the structure of linked pages (the site's information architecture) has a decisive impact on the usability. Previous studies including Shneiderman [29], and Larson and Czerwinski [21] also provided suggestions on how to create the best structure.

Larson and Czerwinski [21] found that users took significantly longer time to find items in a structure with depth than breadth. They compared a three-tiered, eight-links-per-page (8 x 8 x 8) structure with two-tiered, 16 and 32 links per page structures (16 x 32 and 32 x 16).

Bernard [2] had an important contribution to the analysis of hypertext structure. He devised a metric called Hypertext Accessibility Index (HAI) to model the informational accessibility of a particular hypertext structure compared to other alternative structures. The metric was based on the Entropy theory. It explained what Larson and Czerwinski [21] as well as Kiger [20] found about the navigation time of shallow and deep structure in quantitative terms.

Although Bernard's HAI metric is useful for web structure comparison and it considers the level of depth as well as the number of hyperlink at each node, it cannot be used easily in practical because a web site structure also depends heavily on the content and the distribution probability of the information goal. The location and the design of the hyperlink also affect the navigation time.

To consider other factors on the navigation time of the web site, an entropy approach may be employed. Kao et al. [19] proposed the LAMIS method with the entropy analysis to distill the information of a web site. Rather than defining the probability term as the normalized feature frequency in the page set, we may define a probability term for the hyperlink. The probability value will depends on the position, size, menu structure, description, color and the information goal, etc. Empirically, the probability is directly linked to the transition probability of the web log.

A study on the use of entropy theory to merge the web site content was performed by Chen et. al. [6]. The merge depends on the mutual information of term w_i and w_j in the sub-tree as well as their counts. A similar approach will also be applied to our web site structure analysis in order to combine the web site content to reduce the navigation time.

In a study on web site structure, some approaches were developed to restructure the hyperlinks according to the users' browsing behavior. To achieve such goal, collaborative filtering, Markov model, Longest Repeating Subsequences (LRS) [28] or data mining techniques are commonly used. Based on the Markov model, Jenamani et al. [18] proposed several algorithms to examine (1.) the most accessed pages, (2.) the company's interest, (3.) the visitor's interest pages, (4.) the current visitor's interest pages, (5.) the customized index generation algorithms. However, our system objectives are not to provide a dynamic hyperlink structure, we will concentrate on the study of building a static optimal web site structure.

A web site is viewed as a combination of a set of pages and sub sites as shown in Fig.2. The sub-site structure may be much different from its parent web site. For example, in our department, professional short-term training courses under http://www.comp.hkbu.edu.hk/~training/ are stored in a separate directory and follow a different navigation menu structure than the departmental web site. Such design is very common in a medium to large scale web site. Therefore, to analyze a web site structure, sub sites should be isolated and analyzed in its entity.

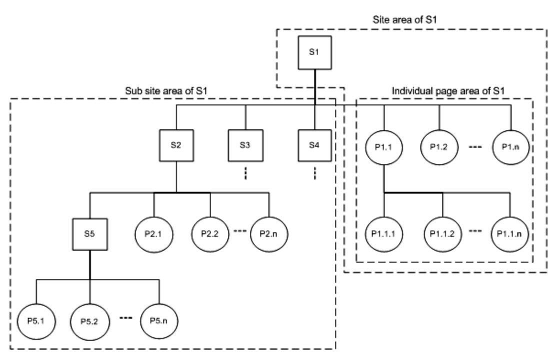


Fig.2 Web Site Structure Model (□ : Site, S1 is the parent web site, S2, S3, etc are the sub sites; ○ : Page, P1.1 refers to the first page in S1, P1.2 is the second)

4. Web Design

Web site layout is composed of many objects, such as menu, buttons, text area, image and graphics, etc. Because the users have their expectations of where web objects are located, the layout design has an influence on the navigation time. Table 1 summarizes what Bernard empirically found in the past few years [1]. These design guidelines will be embedded into the Web Design and Layout Analysis Module of our system.

Web Object	Expected Position
Internal web link	Upper left side
External web link	Right side or lower left side
"Back to home" link	Top-left
Search engine	Top-center
Ad banner	Top
Login / register button	Upper-left
Shopping cart	Top-right
Help	Upper-right
Links to merchandise	Left upper-center
Account / order button	Upper-right
Links with summaries	Most usable
Lists	Best be bulleted
Menu	Index menu accessed faster
Categorical menu	Superior in search performance
Menu links with summary text	More preferable

Table 1 : Web object and expected position

In a study on color usage, Meister and Sullivan [22] reported on the relative legibility of seven colors as a function of symbol size. It was found that white, yellow and red symbols were more easily read than the others. It was also revealed by researchers Shurtleff [30] and Durrett [8] that symbol identification accuracy was best for white and for colors near green and yellow, blue on red was slightly worse.

5. The Webmius Platform

5.1. Task-Based Usability Evaluation Module

The task-based usability evaluation module consists of a front-end web server, a back-end proxy server and a database.

In the front-end, the module has an administration area and a tester area (Fig.3). Inside the administration area, task questions and answers are input by the administrator, a set of target web page locations where the answers can be found are also input to the platform. The target locations are needed to determine whether the user has entered the target answer page before answering the question. If the user does not really enter the target page, the answer will be ignored in the analysis.

In the back-end, user activities are logged into the proxy server which is capable of capturing the query string content in the URL. Therefore, the proxy server may know what dynamic content is requested by the user. However, the access log of the proxy server contains a lot of noise such as image files and multimedia files. To remove such noisy records in the proxy server, the access log is filtered by the module and the filtered access data will be stored in the database.

To start the user test, the platform will send the participant an email which contains a total of 15 tasks, 5 tasks each web site. The tasks are randomly selected from a task pool.

After the user clicks on the corresponding hyperlink of the email, a pop-up window will be opened to start the test. Proxy server configuration is needed to set up in the user's browser. It can be detected by examining whether a designated page requested from the user's browser recorded in our proxy server. As the access log of the proxy server has a record of the designated page, the browser has been configured to connect to the proxy server. If the system cannot find a record in the access log, the browser is not configured properly with connection to our proxy server. In this case the platform will issue a reminder and send help procedures to guide the user to set up a connection to our proxy server.

As the proxy server is configured correctly, the user starts to click on the first task, a new window will be opened to show the tested web site. The user may browse into the web site and find the corresponding answer. After the answer is found, the user has to go back to the task window to input the answer and submit the form. Our platform will compare the proxy record to the user's answer. If the user enters the answer without going to the corresponding webpage, the answer will be recorded but it is marked to indicate the discrepancy. There is a give-up button for each task to let the user skip a task.

After the user completes all tasks, the user has to answer a questionnaire to comment on the web site. The purpose of the questionnaire is to gather the user's comment on the web site. These data will be analyzed later together with the task record to examine any correlation.

The user may log in to the platform to read the statistics of the web evaluation of all participants. In the later stage of the platform development, the user may even suggest a web site for evaluation and provide a corresponding set of questions and answers in the platform.

After all the users have completed the test, the administrator may export the user statistics to a text file for further analysis.

5.2. Web Design & Layout Analysis Module

After all the evaluations are completed, the data will be analyzed using the metrics stated in [17] together with the following metrics :

	Metric	Description
1	Number of hyperlinks	Only visible and internal hyperlinks except the In-Page ones are counted.
2	Hyperlink level	The level that the hyperlink resides is recorded. Level 1 is the root level which is directly visible. If there is pop-up menu or tree menu, level 2 is the first pop-up menu, level 3 is the second pop-up menu.
3	Hyperlink position	• Based on the top left position of the hyperlink • For pop-up menu or tree menu, the hyperlink position is the absolute position measured based on all other hyperlink groups closed but the interested hyperlink group opened.
4	Hyperlink size	• Text – The area covered by the text. (±5 pixels tolerance) • Graph - Image size
5	Word counts	• English : Number of words • Chinese : Number of characters • English and Chinese mix : Number of English words + Number of Chinese characters
6	Word count inside and near the hyperlink	• English : Number of words • Chinese : Number of characters • English and Chinese mix : Number of English words + Number of Chinese characters • For image hyperlink with text content, find the word count based on the text displayed. • For image hyperlink with alt, find the word count inside the alt.

Table 2 : Web Design and Layout Analysis Metrics

For each metric, a weight factor associated with the probability of the click is determined by the combined results of task-based usability evaluation module, the web structure mining module and the semantics inference module.

5.3. Web Structure Mining Module

The web structure mining tool is embedded with a spider which captures the whole web site structure. After downloading the web pages, the hyperlinks of each web page will be indexed. Based on our heuristic hyperlink analyzer, the web page will be segmented into 2 major areas, the hyperlink area and the content area. Other than analyzing with the metrics, the hyperlink type is diagnosed. Seven hyperlink types listed in Table 3 are classified.

	Hyperlink Type	Description
1	External link	Hyperlink pointing to the web site of different subdomains or domains
2	Internal link	
	- In-page link (Self loop) (index 1)	The content and the effect of such hyperlink will be ignored in the analysis
	- Intra-directory links (index 2)	Inside the same directory
	- Up link (index 3)	Point to parent directory
	- Down link (index 4)	Follow the file directory structure, point to the immediate child directory
	- Across links (index 5)	All links within a host that are not of the other types
	- Download link (index 6)	Link for downloading image, file, etc. Regarded as leave

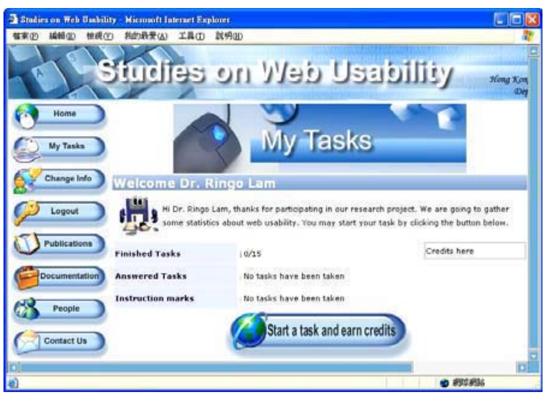
Table 3 : Hyperlink Types

External link will be ignored in further diagnosis because it points to a different web site. On the other hand, the hyperlink type is useful to identify the page characteristics, four page types can be discriminated, they are :

- Home page : is the first page of a set of pages. Its distance to other pages should be small but the number of links should not be large [3]
- Index page : is referred as TOC, i.e. table of contents page, and has a higher number of outlinks.
- Reference page : is like glossary, contains references, and has higher number of inlinks.
- Content page (leave)

The understanding of the page type is fundamental to the web site structure mining. For example, index page is designed for navigation purpose. Information searching should not return the index page.

Besides, the text around the hyperlink is an important metric in analyzing the web page relationship. For example, it was proposed by Chakrabarti [5] to help finding the authority, that is the reference or the content page.



(a.) User test area

(b.) Administration area

Start time	2004-07-21 14:52:09		
End time	2004-07-21 14:54:01		
Datetime	**Index**	**Url**	**Match target**
2004-07-21 14:52:15	1	http://www.comp.hkbu.edu.hk/	
2004-07-21 14:52:15	14	http://www.comp.hkbu.edu.hk/csd_home/index.php	
2004-07-21 14:52:16	14	http://www.comp.hkbu.edu.hk/csd_home/index.php	
2004-07-21 14:52:24	24	http://www.comp.hkbu.edu.hk/csd_home/message_from_head.php	
2004-07-21 14:52:30	21	http://www.comp.hkbu.edu.hk/csd_home/cs_corner.php	
2004-07-21 14:52:32	52	http://www.comp.hkbu.edu.hk/syllabus/syllabus.html	
2004-07-21 14:52:32	2	http://www.comp.hkbu.edu.hk/csd_home/top_banner.php	
2004-07-21 14:52:32	52	http://www.comp.hkbu.edu.hk/syllabus/syllabus.php	
2004-07-21 14:52:35	7	http://www.comp.hkbu.edu.hk/csd_home/education.php	
2004-07-21 14:52:36	36	http://www.comp.hkbu.edu.hk/csd_home/course_is.php	
2004-07-21 14:53:51	20	http://www.comp.hkbu.edu.hk/csd_home/staff.php	
2004-07-21 14:53:52	46	http://www.comp.hkbu.edu.hk/csd_home/staff_academic.php	Match

(c.) Filtered access record

Fig. 3 : Task-Based Usability Evaluation Module

5.4 Semantics Inference Module

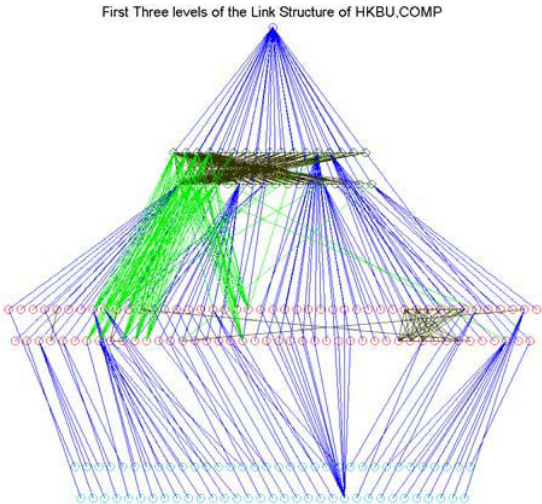
The semantics information from the web page and in particular the descriptive labels for hyperlinks are important for successfully analyzing the web design and usability. We plan to develop word analysis and machine learning tools in this module.

6. Web Design Layout Discovery via Association Rule Mining

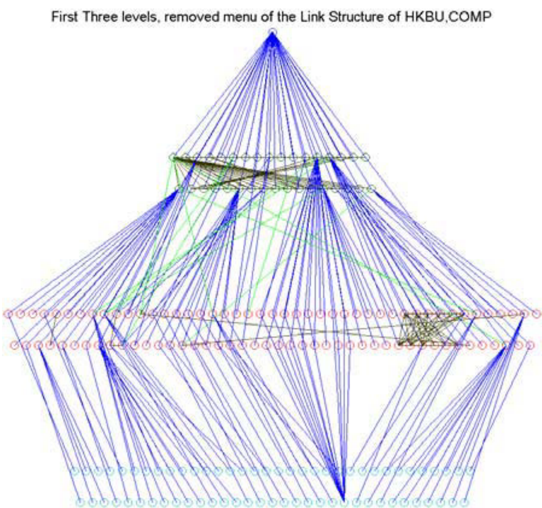
In this section, we will briefly introduce a method of using the web mining for layout discovery. Traditionally, web site design and layout are often recovered manually or aided by authoring tools such as Macromedia Dreamweaver. In this part, we present our results on the recovery of web design layout via Web mining. The visualization of link structures of any non-trivial site is almost impossible due to the large and complex link structures of the sites. Figure 4(a) shows the partial link structure of the first three levels in the web site of Computer Science dept. of HKBU, where the links from the third level to other lower levels are omitted. The dense structure is very difficult to analyze visually and furthermore the links itself are generated from two sources where one type of links originates from the menu structure of the web page and one type of links originates from the actual content of the page.

As the association pattern of links are different in menu links, we propose the use of association rule mining to process the results we obtain in the web structure analysis. In the several web sites we studied, the association rules extraction is very successful in retrieving the top menu system and the sub-site structures of the web sites. Figure 4b shows the link structure after the links originated from menu is removed.

On the other hand, we may find that there are some sub-site structures in our departmental web site. The obvious one is located at the area bounded near 80 – 100 in Fig.5.



(a.) All levels displayed with inlinks and outlinks



(b.) All levels with menu removed and displayed with inlinks, outlinks

Fig. 4 : Link structure of the Computer Science dept. at the Hong Kong Baptist University (Blue line : From the upper to the lower level, Green line : From the lower to the upper level)

The x axis denotes the page with the inlink and the y axis refers to the page with the outlink. There are some horizontal lines in Fig.5. They refer to pages with a number of links pointing to the other pages. In general such pages are usually navigation pages. The vertical lines are usually related to the menu page. Because menu is consistent and almost ubiquitous throughout the web site, there are many pages with links to the menu page.

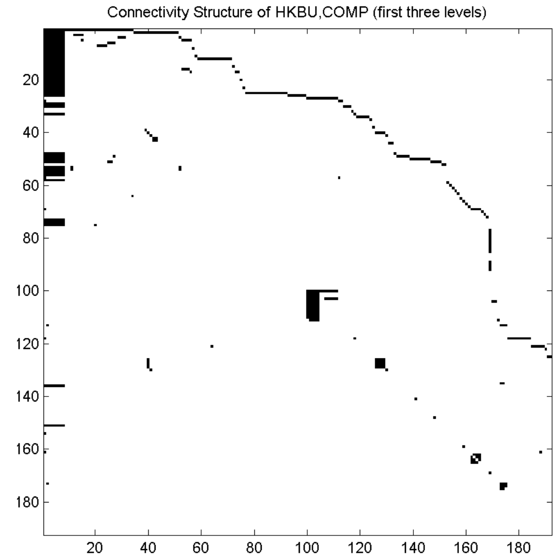


Fig.5 : Transition matrix of our web site

7. Conclusions

We have reviewed the different approaches to web usability study and proposed an integrated web mining and usability analysis system. The system is composed of four modules: the task-based usability evaluation, the web design and layout analysis, the web structure mining and the semantics inference. The task-based usability module has been completed and is being user tested, the web structure mining and design and layout analysis module are partially completed. The preliminary result on web structure mining shows how web mining can provide important design and layout analysis for a web site.

References

[1] Bernard, M. L., “Optimal Web Design”, *http://psychology.wichita.edu/optimalweb/print.htm*

[2] Bernard, M. L., “Examining a Metric for Predicting the Accessibility of Information within Hypertext Structures”, *Dissertation Thesis*, Wichita State University, 2002

[3] Botafogo, R. A., Rivlin, E., Shneiderman, B., “Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics.” *ACM Transactions on Information Systems*, Vol. 10, No. 2, 1992, pp.142 – 180.

[4] Brajnik, G. “Automatic Web Usability Evaluation: What Needs to be Done?”, *Proc. of 6th Human Factors and the Web Conference*, Austin, Texas, June 2000.

[5] Chakrabarti, S., Dom, B., Raghavan, P., Rajagopalan, S., "Automatic Resource Compilation by Analyzing Hyperlink Structure and Associated Text", *Proc of 7the World Wide Web Conference*, 1998

[6] Chen, Z., Liu, S., Liu, W., Pu, G. and Ma, W., "Building a Web Thesaurus From Web Link Structure", *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, 2003, pp.48-55.

[7] Chi, E. H., Rosien, A., Supattanasiri, G., Williams, A., Royer, C., Chow, C, Robles, E., Dalal, B., Chen, J., Cousins, S., "The Bloodhound Project: Automating Discovery of Web Usability Issues Using the InfoScent™ Simulator", *CHI 2003*, Ft. Lauderdale, Florida, USA, April 5–10, 2003.

[8] Durrett, H.J., *Color and the Computer*. Orlando, FL: Academic Press, Inc, 1987.

[9] Glover, E. J. , Tsioutsiouliklis K., Lawrence S., Pennock, D. M. and Flake, G. W., "Using Web Structure for Classifying and Describing Web Pages", *WWW2002*, Honolulu, Hawaii, USA, May 7-11, 2002.

[10] Halkidi, M., Nguyen, B., Varlamis, I. and Vazirgiannis, M., "Thesus: Organising Web Document Collections based on Semantics and Clustering", *Journal on Very Large Databases, Special Edition on the Semantic Web*, November, 2003

[11] Heer, J. and Chi, E. H., "Separating the Swarm: Categorization Methods for User Sessions on the Web", *CHI2002*, Minneapolis, Minnesota, USA, April 20-25, 2002.

[12] Hinckley, K., Cutrell, E., Bathiche, S., and Muss, T., "Quantitative Analysis of Scrolling Techniques", *CHI 2002*, Minneapolis, Minnesota, USA, April 20-25, 2002.

[13] Hong, J. I., Heer, J., Waterson, S., and Landay, J. A., "WebQuilt: A Proxy-based Approach to Remote Web Usability Testing", *ACM Transactions on Information Systems*, Vol.19, Iss.3, 2001, pp.263-385.

[14] Hornof, A. J. and Halverson, T., "Cognitive Strategies and Eye Movements for Searching Hierarchical Computer Displays", *CHI 2003,* Ft. Lauderdale, Florida, USA, April 5-10, 2003.

[15] Ivory, M. Y., Sinha, R. R. and Hearst, M.A., "Preliminary Findings on Quantitative Measures for Distinguishing Highly Rated Information-Centric Web Pages", *Proceedings of the 6th Conference on Human Factors and the Web*, 2000.

[16] Ivory, M. Y. and Hearst, M. A., "The State of the Art in Automating Usability Evaluation of User Interfaces", *ACM Computing Surveys*, Vol.33, No.4, 2001, pp.470 – 516.

[17] Ivory, M. Y., "An Empirical Approach to Automated Web Site Evaluation", *Journal of Digital Information Management*, Vol.1, No.2, 2003, pp.75-102.

[18] Jenamani M., Mohapatra P. K. J., and Ghose S., "Online Customized Index Synthesis in Commercial Web Sites", *IEEE Intelligent Systems*, Vol.17, No.6, 2002, pp.20-26.

[19] Kao, H., Lin, S., Ho, J., Chen, M., "Mining Web Informative Structures and Contents Based on Entropy Analysis", *IEEE Transactions on Knowledge and Data Engineering*, Vol.16, Iss.1, 2004, pp.41 – 55.

[20] Kiger, J. I. , "The Depth / Breadth Tradeoff in the Design of Menu-Driven Interfaces.", *International Journal of Man-Machine Studies*, Vol.20, 1984, pp.201-213.

[21] Larson, K., & Czerwinski, M. , "Web page design: Implications of memory, structure and scent for information retrieval", *CHI'98: Human Factors in Computing Systems*, New York: ACM Press, 1998, pp.25-32.

[22] Meister, D., & Sullivan, D.J. "Guide to Human Engineering Design for Visual Displays" Bunker-Ramo Corp., 1969

[23] McCarthy, J, Sasse, M.A. & Riegelsberger, J., "Could I Have the Menu Please? An eye Tracking Study of Design Conventions", *Proceedings of HCI2003,* Bath, UK, Sep 8-12,, 2003.

[24] Miller, C. S. and Remington, R. W., "Modeling Information Navigation : Implications for Information Architecture", *Human-Computer Interaction*, Vol.19, No.3, 2004.

[25] National Institute of Standards and Technology, "Overview WebSAT", http://zing.ncsl.nist.gov /WebTools/WebSAT/overview.htm, 2001.

[26] Obendorf, H., Weinreich, H., Hass, T., "Automatic Support for Web User Studies With SCONE and TEA", *CHI 2004,* Vienna, Austria, ACM, April 24–29, 2004.

[27] Paganelli, L and Paterno, F., "Automatic Reconstruction of the Underlying Interaction Design of Web Applications", *Proceedings of the 14th international conference on Software engineering and knowledge engineering*, 2002, pp.439 – 445.

[28] Pitkow, J. E. and Pirolli, P. "Mining Longest Repeated Subsequences to Predict World Wide Web Surfing.", *Second USENIX Symposium on Internet Technologies and System, 1999.*

[29] Shneiderman, B., "Designing the User Interface", Strategies for Effective Human-Computer Interaction, 3rd ed, Reading, MA : Addison-Wesley, 1998.

[30] Shurtleff, D.A., *How to Make Displays Legible*, La Mirada, CA: Human Interface Design, 1980.