

Association for Information Systems AIS Electronic Library (AISeL)

PACIS 2004 Proceedings

Pacific Asia Conference on Information Systems
(PACIS)

December 2004

A Comparative Analysis of Automated Web Site Evaluation Tools

Youwei Wang
Fudan University

Dall Kim
Dankook University

Follow this and additional works at: <http://aisel.aisnet.org/pacis2004>

Recommended Citation

Wang, Youwei and Kim, Dall, "A Comparative Analysis of Automated Web Site Evaluation Tools" (2004). *PACIS 2004 Proceedings*. 141.
<http://aisel.aisnet.org/pacis2004/141>

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2004 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

A Comparative Analysis of Automated Web Site Evaluation Tools

JaeHwa Choi
Dankook University
jchoi@dankook.ac.kr

DallYong Kim
Dankook University
dallyong@dankook.ac.kr

Abstract

Web sites have become a primary means of commerce and information exchange. Majority of web sites are designed by non-experts. Although there are guidelines and tips for designing a good web site, usability and accessibility remain critical issues to be addressed by non-experts. Automated tools have emerged to help solve the problem. This paper reviews web site evaluation researches and performs a comparative analysis on major web site evaluation tools for the research and commercial use.

Keywords: Automated Web Site Evaluation, Usability, Accessibility

I. Introduction

The number of web sites increases continuously and new pages are added to the existing web site. Nielsen(2000) predicts that the number of web pages will reach 50 billion and the number of web site increases to 200 million in 2005. Everyday new web sites are created and existing web sites are changed. The creation and change of a web site is designed, developed, and operated by non-experts rather than experts. Therefore, usability and accessibility of a web site become a critical issue although there are guidelines and tips for an effective web site development.

Non-experts, who are not educated and trained for the web design formally, often use design guidelines and web site evaluation tools which are developed by experts through their experience and the user survey. It is not an easy task for a non-expert web designer to study carefully and apply guidelines in designing web sites. In general individual guidelines often conflict with each other and the same guideline is suggested for all kinds of web sites regardless of their objectives. Sometimes guidelines are not realistic. The advance in web technologies creates new guidelines. So it is not easy to design a high quality web site using guidelines only, the importance of a web site evaluation tool is great for non-experts.

Today web sites become the primary means of the commerce and information exchange. In particular, corporate web sites may have a critical impact on the success of the internet business market(Song and Zahedi, 2001). Thus it is important to study the design methodology in order to increase web usability and accessibility. An automated web site evaluation tool is one of the solutions to address the problem. Currently a number of automated web site evaluation tools have been introduced.

This paper summarizes the web site evaluation theories and conducts a comparative analysis on the automated web site evaluation tools. The structure of the paper is as following: Chapter 2 summarizes previous studies on the web site evaluation and the web site evaluation tool. Chapter 3 describes the web site evaluation and reviews the function and types of web site evaluation tools. Chapter 4 proposes a model to analyze web site evaluation tools and discuss the result of comparison of web site evaluation tools using the model. Chapter 5 concludes the research with a direction for future work.

II. Literature Review

2.1 Web Site Evaluation Studies

The web site evaluation may be approached from two different perspectives: user and web site designer/administrator. From the user's perspective on the web site evaluation, most studies focus on the factors for successful web sites. These researches concentrate on the development of a web site evaluation tool. These studies search for design and content elements of a successful web site using the exploratory study.

Olsina et al.(1999) suggests function, usability, efficiency, and reliability as main criteria for the web site quality evaluation. Olsina et al.(1999) tests their QEM(Web-site Quality Evaluation Method) with 6 university sites from different countries. Huizingh(2000) classifies the web site architecture into content and design and specifies each category into evaluation criteria according to the characteristics and perception of a web site. They test the framework with web sites in Yahoo and Dutch yellow page and summarize the findings based on the industry and the size of the web site.

Mateos et al.(2001) developed a web site evaluation model to test university web sites in Spain. The model, called Web Assessment Index(WAI), has content, accessibility, navigation at the major criteria. Palmer(2002) developed metrics for web site usability, design, and performance and conducted a user test with them. Through three consecutive tests Palmer(2002) concluded that the success of a web site is dependent on the speed, navigation, content, interactivity, and response.

From the web site designer or administrator's perspective the web site evaluation focuses on the web usability and accessibility. The web site evaluation model is based on the study of the user-centered development and evaluation approach. This study attempts to develop the methodology and tool for the web site quality evaluation from the information systems and software engineering perspective.

Sinha, R. et al(2001) and Ivory, M. Y. et al(2002b) investigated best web sites selected by experts and users in order to identify the common characteristics of them. To empirically determine whether content is more important than graphics, Sinha, R. et al(2001) examined Webby Award 2000 data set to distinguish the factors of best web sites from the factors of other web sites. Webby Award evaluators use 5 specific criteria and the general experience. The criteria include content, structure and navigation, visual design, functionality, and interactivity. Although content was found to be more important than graphics, Sinha, R. et al(2001) concludes that evaluation criteria can not be considered independently. Ivory, M. Y. et al(2002b) confirmed that their 154 criteria had high accuracy rate of 84% by applying them to 157 web pages and Webby Award 2000 data set.

2.2 Web Site Evaluation Tool Studies

Ivory and Hearst(2001) give a comprehensive review on the automation of user interface usability evaluation including automated web site evaluation tools. In this survey, Ivory and Hearst summarize the usability evaluation methods and propose a new methodology. This new methodology, called WebTango, is introduced in Ivory and Hearst(2002). The WebTango is a quality checker, which aims to help nonprofessional designers improve their sites using quantitative measures of the informational, navigational, and graphical aspects of a Web site.

Brajnik(2000) applies the usability evaluation approach in the field of the software engineering to the web site usability evaluation. Brajnik compares automated evaluation tools

using consistency, adequate feedback, situational navigation, efficient navigation, and flexibility as the characteristics of usability.

Schubert and Selz(1999) propose a web site evaluation model based on the stages of a transaction in the electronic market. They identify three stages of the electronic commerce - information stage, contract stage, and payment stage – and assume that the communication stage in the cyber community plays an important role. Their web site evaluation model is based on the stages and was applied to the Swissair web site. Brajnik(2002) develops a web site evaluation model by applying the software quality model. Brajnik(2002) proposes a test method to determine whether an automated web site evaluation tool uses the proper rules and applies it to the LIFT, an automated web site evaluation tool. Ivory et el(2002b) verifies the validity of a set of web site evaluation criteria using the Webby Award 2000 data set.

III. Web Site Evaluation Tool

3.1 Web Site Evaluation

The web site designer or administrator evaluates a web site design during the web site life cycle so that the web site becomes a successful one. In general, the web site evaluation can be done through either the preliminary review, conformance evaluation, or ongoing monitoring. Preliminary review identifies general problems of a web site, conformance evaluation finds major violations of guidelines during the web site design stage. Conformance evaluation generally checks which level of Web Content Accessibility Guidelines(WCAG; W3C WCAG 2.0) a web site satisfies. Ongoing monitoring tries to make sure that a web site maintains a certain level of WCAG.

There are two methods of web site evaluation: automated evaluation and non-automated evaluation. Non-automated web site evaluation can be done in two ways. The first approach is user testing. This approach allows web site user groups to evaluate a web site and collects opinions about evaluation criteria and analyzes them. The second approach is a heuristic testing. This method asks experts to identify factors which will affect to the web site users.

Heuristic testing costs high since it asks experts to conduct an analysis and develop reports after the analysis. Thus, it is difficult for many organizations to hire web site experts for the regular evaluation. User testing requires to make the evaluation criteria and environment very clear. To measure subjective features like usability, user testing requires to define a standard procedure to produce a repetitive and comparable result.

For most web sites neither heuristic testing nor user testing is not practical due to two reasons. First, rapid advance in the web technology makes the use of sophisticated tools and complex interaction of a web site possible. Second, the life cycle of a web site is very short. The web site improvement has to be done faster than other software maintenance due to market pressure and the lack of barrier in web site development. These characteristics of the web site evaluation methods make an automated web site evaluation tool a necessity, not an option.

Automated web site evaluation tools play a bigger role in supplementing or substituting non-automated web site evaluation tool. Automated web site evaluation tools allow to identify potential usability problems before the actual operation of a web site and to select the best design through the comparison of alternative designs. This reduces economic and non-economic cost of non-automated web site evaluations. Automated web site evaluation tools also allow the web site designer or administrator to evaluate many web sites and to detect potential problems as well as actual problems.

3.2 Automated Web Site Evaluation Tool

The function of an automated web site evaluation tool largely consists of capture, analysis, critique of web site data(Ivory et el, 2001). Capture activity records usage data. Analysis activity identifies potential usability problems. Critique activity proposes improvements for potential problems.

Web Accessibility Initiative(WAI) of W3C classifies automated web site evaluation tools into evaluation tool, repair tool, and transformation tool. Ivory et el(2001) divides the analysis tools of automated web site tools into four types as shown in Table 1:

<Table 1> Types of Automated Web site evaluation tool

| Type | Analyze site usage data | Check guideline conformance | Simulate hypothetical user(s) navigating a site | Check server performance |
|--------------|---|--|---|---|
| Synop-sis | ◆analyze server log file data to identify potential problems in usage patterns | ◆detect and flag a web page or site's deviation from design guidelines | ◆minic hypothetical user's browsing experience and output computed measures | ◆Monitor consistency, availability, performance of a web server |
| Advan-tages | ◆examine trends in site usage, errors, clickstreams, etc | ◆objective, consistent assessment ◆reduce designer overload | ◆simulate broad use of a site | ◆assess & react to system performance once site goes live ◆identify potential performance problems before going live |
| Limita-tions | ◆ cannot assess user intentions or satisfaction with a site ◆only as good as the logged data | ◆cannot assess user satisfaction with a site ◆cannot automatically detect conformance to all guidelines | ◆cannot assess user satisfaction with a site ◆only as good as the simulation model | ◆cannot assess user satisfaction with a site |
| Tools | ◆WebTrends Reporting Center ◆ WebQuilt ♣ CAST's Bobby | ◆Web Static Analyzer Tool, W3C HTML Validator ◆WebTango ◆Faraday's Design Advisor | ◆Web AIM Cognitive Disability Simulation ◆ColiDes Web Navigation Demo | ◆Resource Analyzer, ◆Exodus Performance Monitoring |

There are four types of automated web site evaluation tools which identify potential usability problems of a web site. The first type of tools analyzes server log file data to identify potential problems in usage patterns. The second type of tools help check whether the HTML code of a web site follows the proper coding practice from a usability point of view. The third type of tools evaluates a web site's usability by collecting data through a simulation of a hypothetical user's experience. The fourth type of tools monitors consistency, availability, and performance of a Web server by stressing the server. The second type of tools is most widely

used in practice and some of the examples includes A-Prompt, WatchFire Bobby, UsableNet LIFT, W3C HTML Validator, and NIST(National Institute of Standards and Technology) WebSAT(Web Static Analyzer Tool).

A-Prompt, WatchFire Bobby, UsableNet LIFT, W3C HTML Validator, and NIST WebSAT examine HTML to evaluate a web site's usability. These tools check the conformance of WCAG or Section 508 guidelines. In 1998 U.S. government, the federal law Rehabilitation Act 508, requires all electronic information technologies allow handicap people to use them. Therefore every web site is required to provide assessability to all and this guideline becomes an evaluation criteria of automated web site evaluation tools.

Max of WebCriteria evaluates the usability of a web site by collecting primary statistical data through the simulation model. The primary evaluation criteria include assessability, load time, and content. NetRaker, another automated web site evaluation tool, develops an online survey which allows users to answer the survey while using the web site. NetRaker does not check HTML code or analyze statistical data. Instead, it collects and analyzes user survey data of a web site.

IV. Comparison of Automated Web Site Evaluation Tools

4.1 A Model for the Web Site Evaluation Tool Comparison

Web site designer or administrator evaluates the web site to make it successful during the life cycle of a web site. The web site evaluation model helps the web site designer or administrator understand, control, and improve the web site and the development process. The web site evaluation model defines the standardized measurement and data collection which will allow to develop a high quality web site. Thus, the web site evaluation model becomes the criteria to measure the quality level of a web site.

Web site design guidelines may be considered as a web site evaluation model. Examples include the WCAG and the Nielsen guideline(Nielsen 2000). The Web Accessibility Initiative(WAI) is an effort by the W3C organization to improve website accessibility. They publish a set of guidelines(W3C WCAG 2.0) where accessibility is defined as the website ability to be used by someone with disabilities. An accessible website ensures graceful transformation and makes content understandable and navigable. And the guidelines of established web site evaluation agencies are another example since they influence the web site designer and administrator. There are other informal guidelines suggested by experienced web site designers.

Brajnik(2002) approaches the web site evaluation from a web site quality perspective. Brajnik(2002) does not propose a specific web site evaluation model although it insists that a web site evaluation model needs to be developed with a software quality concept. Yet the quality approach to a web site can be applied to understand, control, and improve the development process and its result during a web site design.

From a product quality dimension, the primary purpose of a web site evaluation is to evaluate a web site's usability. Usability of a web site is about the ease of use of a web site. ISO 9241-11, the international quality standard, defines the usability as the level of effectiveness, efficiency, and satisfaction of a user who wants to achieve its purpose using the product in a specific environment.

In order to be operationalized the accessibility and usability properties need to be decomposed into more detailed ones. All these properties may be further decomposed into more detailed ones that refer to specific attributes of the website implementation. Brajnik(2004) distinguishes internal attributes from external attributes. Internal attributes

depend only on how the website has been designed/developed, while external attributes depend on the website and its usage. While for evaluating usability of a website both internal and external attributes are needed, only the former ones are amenable for automatic tests. External attributes can be evaluated only via semi-automatic means that entail a human evaluation step(Brajnik, 2004).

We developed a web site evaluation model as shown in Figure 1 to compare automated web site evaluation tools. To check whether an automated web site evaluation tool addresses both the accessibility and the usability of a web site, the proposed model has criteria representing both the accessibility and the usability.

The web site evaluation criteria at the highest level are classified into efficiency and effectiveness. In this model satisfaction is not considered as a separate criterion since it will be achieved through effectiveness and efficiency. Usability is to give satisfaction through effectiveness and efficiency.

The top level evaluation criteria are further divided into consistency, navigation, maintainability, accessibility, and performance at the next level. The criteria at the second level are consistent with the criteria used in other studies. Accessibility is included as usability criteria since usability implies accessibility, not vice versa. Specific criteria for the web site evaluation are not mutually exclusive nor independent. Thus one criterion may be related to two or more attributes as shown in Figure 1.

4.2 Comparison of Automated Web Site Evaluation Tool

This study applies a web evaluation model shown in Figure 1 to 5 popular automated web site evaluation tools: A-Prompt, UsableNet LIFT, Watchfire Bobby, NIST WebSAT, and AnyBrowser. These tools evaluated a university web site and produced a result shown in Table 2.

Among the tools tested A-Prompt can analyze the usability of a web site and modify it, other tools can only analyze a web site. A-Prompt detects the problem of a web site and allows to correct the HTML source code using A-Prompt. Other tools detect the problem of a web site and suggest the guideline to modify it.

Among the 5 automated web site evaluation tools tested, A-Prompt, UsableNet LIFT, and Watchfire Bobby are tools which use WCAG and Section 508 as their basic guidelines, NIST WebSAT uses its own usability guideline to evaluate a web site. AnyBrowser not only checks the general guideline, but also evaluates the usability of a web site based on the resolution, the brand and version of a web browser, and the HTML version.

Except the NIST WebSAT which uses its own usability guideline, other tools test primarily the accessibility items. WebSAT(Web Static Analyzer Tool) examines HTML of a web site and identifies its usability into five categories: Accessibility, Form Use, Performance, Maintainability, Navigation, Readability. Other tools don't correct potential problems of a web site usability which are tagged with 'M' in Table 2, and ask evaluators to manually check them with general guidelines. This is typical with other automated web site evaluation tools which are not tested in this study.

The reason that most tools are concentrated on the accessibility is that tools are designed to check WCAG and Section 508 guidelines. To test the overall usability of a web site tools need to develop their own guidelines like NIST WebSAT. To go beyond providing the suggestion of manual check of a web site, automated web site evaluation tools need to be improved to conduct intelligent analysis. Automated web site evaluation tools will become more intelligent with the help of the research in cognitive science and user interface

V. Conclusion

This study reviews the previous work on the web site evaluation and automated web site evaluation tool and reports the result of a comparative study of 5 automated web site evaluation tools using a web site quality evaluation model. Automated web site evaluation tools can complement direct evaluation with non-automatic tools in order to improve the usability of a web site. Current automated web site evaluation tools need to be improved so that they can do beyond the typical design guideline check. Thus, future research on the automated web site evaluation tool should concentrate on providing a specific direction and method to analyze the usability of a web site.

References

- Alexander, J. and Tate, M. A., Web Wisdom: How to Evaluate and Create Information Quality on the Web, Lawrence Erlbaum Associates, 1999
- Brajnik, G., "Automatic Web Usability Evaluation," Proc. of Human Factors and the Web, 6th Conference, June 2000
- Brajnik, G., "Quality Models based on Automatic Webtesting," CHI 2002 Workshop, 2002
- Brajnik, G., "Using automatic tools in accessibility and usability assurance," Lecture Notes in Computer Science, Proc. of the 8th ERCIM UI4ALL Workshop, C. Stephanidis ed., Springer Verlag, Vienna, June 2004
- Bauer, C. and Schal, A., "Quantitive evaluation of Web site content and structure," Internet Research: Electronic Networking Applications and Policy, Vol. 10, No. 1, 2000
- Huizinth, Eelko K.R.E., "The Content and Design of Web Sites: An Empirical Study," Information & Management, Vol.37, 2000
- Ivory, M. Y. and Hearst, M. A., "The State of the Art in Automating Usability Evaluation of User Interfaces," ACM Computing Surveys, 2001, December
- Ivory, M. Y. and Hearst, M. A., "Improving Web Site Design," IEEE Internet Computing, 2002a, March/April
- Ivory, M. Y. and Hearst, M. A., "Statistical Profiles of Highly-Rated Web Sites," CHI 2002b, April
- Lynch, P. J. and Horton, S., Web Style Guide: Basic Design Principles for Creating Web Sites, Yale University Press, 1999
- Mateos, M. B., Mera, A. C., Gonzalez, F. J. M., and Lopez, O. R. G., "A new Web assessment index: Spanish university analysis," Internet Research: Electronic Networking Applications and Policy, Vol. 11, No. 3, 2001
- Nielsen, J., Designing Web Usability, New Riders Publishing, Indianapolis, IN, 2000
- Olsina, L., Godoy, D., Lafuente, G. J. and Rossi, G., "Specifying quality characteristics and attributes for Websites," First ICSE Workshop on Web Engineering(WebE-99), Los Angeles, USA, 1999
- Palmer, W. P., "Web Site Usability, Design, and Performance Metrics," Information Systems Research, Vol.13, No.2, June 2002
- Sano, Darrell, Designing Large-scale Web site: A Visual Design Methodology, Wiley Computer Publishing, 1996
- Schubert, P. and Selz, D., "Web Assessment – Measuring the Effectiveness of Electronic Commerce Sites Going Beyond Traditional Marketing Paradigms," Proc. of 32nd HICSS, 1999
- Sinha, R., Hearst, M., and Ivory, M., "Content or Graphics?: An Empirical Analysis of

Criteria for Award-Winning Websites,” Proceedings of the 7th Conference on Human Factors & the Web, June 2001

Song, J. and Zahedi, F. M., “Web Design in E-Commerce: A Theory and Empirical Analysis,” Proc. of 22nd ICIS, 2001

W3C Web Content Accessibility Guideline(WCAG) 2.0:
<http://www.w3.org/TR/2004/WD-WCAG20-20040311/>

U.S. Section 508 Guidelines: <http://www.access-board.gov/sec508/508standards.htm>

A- Prompt: <http://aprompt.snow.utoronto.ca/>

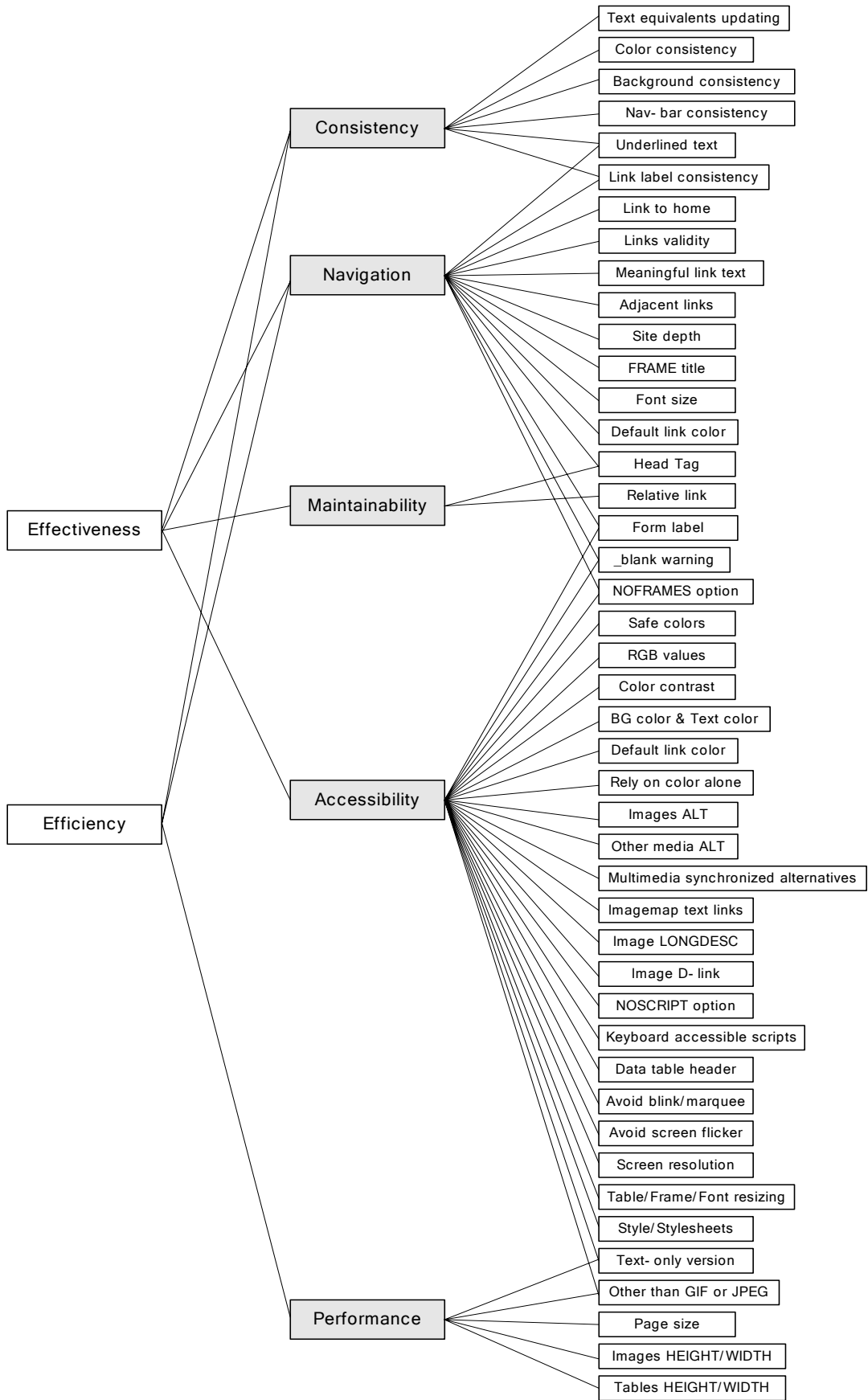
UsableNet LIFT: <http://www.usablenet.com/>

Watchfire Bobby: <http://bobby.watchfire.com/bobby/html/en/index.jsp>

NIST WebSAT: <http://zing.ncsl.nist.gov/WebTools/WebSAT/overview.html>

AnyBrowser HTML Validator: <http://www.anybrowser.com/validateit.html>

[Figure 1] Web Site Evaluation Model



<Table 2> Test Criteria of Automated Web Site Evaluation Tools

| TEST | A-Prompt | UsableNet LIFT | Watchfire Bobby | NIST WebSAT | AnyBrowse r |
|--------------------------------------|----------|-------------------|--------------------|----------------|----------------|
| Text equivalents updating | M | | | | |
| Color consistency | | | | | ○ |
| Background consistency | | | | | |
| Nav-bar consistency | | | M | | |
| Underlined text | | | | | |
| Link label consistency | | | | | |
| Form label | ○ | | ○ | ○ | |
| Link to home | | | | | |
| Links validity | | | | | ○ |
| Meaningful link text | ○ | | M | | |
| Adjacent links | | | ○ | ○ | |
| Site depth | | | | | |
| Frame title | ○ | ○ | ○ | | |
| Font size | | | | | |
| Default link color | | | | ○ | |
| Head tag | | | | ○ | |
| Relative link | | | ○ | ○ | |
| _blank warning | ○ | | ○ | ○ | |
| NOFRAMES option | ○ | ○ | | ○ | |
| Safe colors | ○ | | | | ○ |
| RGB values | | | | ○ | |
| Color contrast | ○ | M | ○ | | |
| BG color & Text color | | | | ○ | |
| Default link color | | | | | |
| Rely on color alone | M | M | | | |
| Images ALT | ○ | ○ | ○ | ○ | |
| Other media ALT | ○ | ○ | ○ | ○ | |
| Multimedia synchronized alternatives | M | ○ | ○ | | |
| Imagemap text links | ○ | ○ | | ○ | |
| Image LONGDESC | ○ | ○ | | | |
| Image D-link | ○ | ○ | | | |
| NOSCRIP T option | ○ | ○ | M | | |
| Keyboard accessible scripts | ○ | ○ | ○ | | |
| Data table header | ○ | ○ | ○ | | |
| Avoid Blink/Marquee | ○ | ○ | ○ | | |
| Avoid Flicker | M | ○ | M | | |
| Screen resolution | | | | | ○ |
| Table/Frame/Font resizing | | | | | |
| Style/Stylesheets | M | ○ | M | | |
| Text-only version | | M | M | | |
| Other than GIF or JPEG | | | | ○ | |
| Page size | | | | ○ | |
| Images HEIGHT/WIDTH | | | | ○ | |
| Tables HEIGHT/WIDTH | | | | | |

* Remark: M - Manually check