International Journal of Engineering Research and Development
e-ISSN: 2278-067X, **p**-ISSN: 2278-800X, www.ijerd.com
Volume 8, Issue 1 (July 2013), PP. 34-37

A Review of Trends in Research on Web Accessibility:

Sushree Sudesna Ashe, Subhendu kumar Pani
Associate Prof,Dept. of CSE OEC,Odisha

Abstract:- In recent years the growth of the World Wide Web exceeded all expectations.The World Wide Web is a fertile area for data mining research.Web mining is a research topic which combines two of the activated research areas: Data Mining and World Wide Web. Web mining research relates to several research communities such as Database, information Retrieval and Artificial intelligence, visualization. Automated analyses of WCAG 2.0 Level,a success criteria found high percentages of violations overall. Unlike more circumscribed studies, however, the sites exhibited improvements over the years on a number of accessibility indicators, with government sites being less likely than top sites to have accessibility violations. Examination of the causes of success and failure suggests that improving accessibility may be due, in part, to changes in website technologies and coding practices rather than a focus on accessibility per se. This paper reviews the research and application issues in web Accessibility besides proving an overall view of Web mining.

Keywords:- Web, Data mining, web usage mining, web content mining, web structure mining,web accessibility.

I. INTRODUCTION

As the economics of service and information provision drive more content exclusively to the Web, a large number of people with disabilities are disadvantaged. Once considered a luxury, access to the Web is now becoming a requirement for full participation in society. Internet is the shared global computing network. It enables global communications between all connected computing devices. It provides the platform for web services and the World Wide Web[7,8,9]. Web is the totality of web pages stored on web servers. There is a spectacular growth in web based information sources and services. It is estimated that, there is approximately doubling of web pages each year. As the Web grows grander and more diverse, search engines also have assumed a central role in the World Wide Web's infrastructure as its scale and impact have escalated. In Internet data are highly unstructured which makes it extremely difficult to search and retrieve valuable information. Search engines define content by keywords. With the explosive growth of information sources available on the World Wide Web, it has become increasingly necessary for users to utilize automated tools in order to find, extract, filter, and evaluate the desired information and resources. In addition, with the transformation of the Web into the primary tool for electronic commerce, it is imperative for organizations and companies, who have invested millions in Internet and intranet technologies, to track and analyze user access patterns. These factors give rise to the necessity of creating server-side and client-side intelligent systems that can effectively mine for knowledge both across the Internet and in particular Web localities. Many organizations and corporations

Provide information and services on the web such as automated customer support, on-line shopping, and a myriad of resources and applications.web based applications and environments for electronic commerce, distance education, on-line collaboration, news broadcasts etc., are becoming common practice and widespread. The WWW is becoming ubiquitous and an ordinary tool for everyday activities of common people, from a child sharing music files with friends to a senior receiving photographs and messages from grandchildren across the world. It is typical to see web pages for courses in all fields taught at universities and colleges providing course and related resources even if these courses are delivered in traditional classrooms. It is not surprising that the web is the means of choice to architect modern advanced distance education systems. There are several important issues, unique to the Web paradigm that comes into play if sophisticated types of analyses are to be done on server side data collections. These include the necessity of integrating various data sources such as server access logs, user registration or profile information; resolving difficulties in the identification of users due to missing unique key attributes in collected data; and the importance of identifying user sessions or transactions from usage data, site topologies, and models of user behavior. We devote the main part of this paper to the discussion of issues and problems that characterize Web Accessibility. Furthermore, we survey some emerging tools and techniques, and identify several future research directions[6,10].

This paper has been organized as follows. The next section presents an overview of classification of web mining. Techniques on the web mining are discussed in section 3. Section 4 discusses success criteria of web Accessibility. Section 5 concludes the paper.

II. WEB DATA MINING

2.1 OVERVIEW

The web mining is the use of data mining techniques to automatically discover and extract information from World Wide Web documents and services [5]. This area of research is so huge today partly due to the interest in e-commerce. This phenomenon partly creates confusion what constitutes Web mining and when comparing research in this area. Similar to [5], we suggest decomposing Web mining into these subtasks, namely

1. Resource finding: the task of retrieving intended Web documents.
2. Information selection and pre-processing: automatically selecting and pre-processing specific information from retrieved Web resources[1].
3. Generalization: automatically discovers general patterns at individual Web sites as well as across multiple sites.
4. Analysis: Validations and/or interpretation of the mined patterns.

We should also note that humans play an important role in the information or knowledge discovery process on the web since the web is an interactive medium. This is especially important for validation and/or interpretation in step 4.So, interactive querytriggered knowledge discovery is as important as the more automatic data triggered knowledge discovery. However, we exclude the knowledge discovery done manually by humans. Thus, Web mining refers to the overall process of discovering potentially useful and previously unknown information or knowledge from the web data. It implicitly covers the standard process of knowledge discovery in databases (KDD) [2].We could simply view web mining as an extension of KDD that is applied on the Web data. From the KDD point of view, the information and knowledge terms are interchangeable[3].There is a close relationship between data mining, machine learning and advanced data analysis[4].Web mining is often associated with IR or IE.

III. WEB MINING TECHNIQUES

Traditional data mining techniques can also be used for web mining, such as classification, clustering, association rule mining, and visualization. In web mining, classification algorithms can be used to classify users into different classes according to their browsing behavior, for example according to their browsing time. After classification, a useful classification rule like "30% of users browse product/food during the hours 8:00-10:00 PM" can be discovered. The difference between classification and clustering is that the classes in classification are predefined (supervised), but in clustering are not predefined (unsupervised). The criterion by which items are assigned to different clusters is the degree of similarity among them. The main purpose of Clustering is to maximize both the similarity of the items in a cluster and the difference between clusters [12].The association rule technique can be used to indicate pages that are most often referenced together and to discover the direct or indirect relationships between web pages in users' browsing behavior [11]. For example, an association rule in the web usage mining area could take the form "the people who view web page index.htm and also view product.htm the support=50% and the confidence=60%". Visualization is a special analytical technique in web mining that allows data and information to be understood or recognized by human eyes by using graphical and visualized means to represent data, information and analysis results [13].In web structure mining, it usually plays an important role in illustrating the structure of hypertexts and links in a website or the linking relationship between websites. For the other two types of web mining technique, visualization is also an ideal tool to model the data or information. For example, a graph (or map) can be used for web usage mining to present the traversal paths of users or a graph may show information about web usage. This approach enables the analyst to understand and efficiently interpret the results of web usage mining.

Association Rules: After transactions are detected in the preprocessing phase, frequent item-sets are discovered using the A-priori algorithm [e.g. 13]. The support of item-set I is defined as the fraction of transactions that contain I and is denoted by $\sigma(I)$. A hypergraph is an extension of a graph where each hyperedge can connect more than two vertices. A hyperedge connects URLs within a frequent item-set. Each hyperedge is weighted by the averaged confidence of all the possible association rules formed on the basis of the frequent item-set that the hyperedge represents. The hyperedge weight can be perceived as a degree of similarity between URLs (vertices).

Sequential Pattern: Patterns are used to discover frequent subsequences among large amount of sequential data. In web usage mining, sequential patterns are exploited to find sequential navigation patterns that appear in users sessions frequently. The typical sequential pattern has the form[15]:the 70% of users who first visited A.html and then visited B.html afterwards ,in the same session,have also accessed page C.html.Sequential patterns might appear syntactically similar to association pattern mining.

Clustering: Clustering: techniques look for groups of similar items among large amount of data based on a general idea of distance function which computes the similarity between groups.Clustering has been widely used in Web Usage Mining to group together similar sessions [14]. Besides information from Web log files,

customer profiles often need to be obtained from an on-line survey form when the transaction occurs. For example, you may be asked to answer the questions like age, gender, email account, mailing address, hobbies, etc. Those data will be stored in the company's customer profile database, and will be used for future data mining purpose.

IV. SUCCESS CRITERIA EVALUATION

Analyses will be grouped within each of the four WCAG 2.0 principles: Perceivability (WCAG 2.0 SC 1.x), Operability (WCAG 2.0 SC 2.x), Understandability (WCAG 2.0 SC 3.x), and Robustness (WCAG 2.0 SC4.x). The relationship to WCAG 1.0 will be noted in each case.

4.1. Perceivability

The perceivability principle addresses the fact that Web users with vision or hearing loss will be unable to access some forms of content unless it can be changed into other perceivable forms.

4.1.1. Text Alternatives: Images. Arguably the most researched aspect of the WCAG guidelines is the inclusion of alternative descriptions of images, often referred to as the "alt tag", such as the following.

Complex images may also include a long description of their content using the longdesc attribute, such as what follows.

.

4.2. Operable

The WCAG 2.0 principle for operability addresses the need for Web pages to provide sufficient time for input to be completed by users of assistive technologies, to avoiding flashing that could cause seizures, and to be navigable by people using diverse means (e.g., a keyboard instead of a mouse). In our automatic checking, we were able to reliably test two of the success criteria related to navigation.

4.3. Understandable

The WCAG 2.0 principle for understandability is designed to ensure that page content and interface controls are understandable to all users. The different aspects of this address issues that affect page readability, the predictability of page appearance and operation, and issues that affect users' ability to avoid and recover from errors.

4.4. Robustness

Robustness refers to the fact that the HTML/XHTML code must be robust enough that assistive technologies (user agents) can reliably interpret it. If the code is malformed, user agents will find it difficult to accurately render pages.

V. CONCLUSIONS

Looking over the entire 14 years since WCAG 1.0, however, we see evidence of improvement, at least for some success criteria, with government sites showing more improvement than top sites on some, but not all, criteria.A more detailed examination of both successes and failures leads us to question the extent to which accessibility gains can be attributed to awareness of and adherence to WCAG guidelines, however. At least some improvements appear to be side-effects of good coding practices and the desire to improve Web page design and website prominence in search results. Future improvements in accessibility may come from explicitly designing new technologies to have such beneficial side-effects.

REFERENCES

[1] G.Salton and M.McGill .introduction to modern information Retrival. McGraw Hill, 1983.

[2] U.Fayyad, G.Piatetsky-Shapiro, P.Smythfrom data mining to knowledge discovery: An overview. In advances in knowledge Discovery and data mining, pages 1-34.AAA Press, 1996.

[3] U.Fayyad, G.Piatetsky-Shapiro, and P.Smyth knowledge discovery and data mining: toward a unifying framework. In proceeding of the second int. conference on Knowledge Discovery and Data mining, pages 82-88, 1996.

[4] M.A.Hearst.Untangling text data mining. In proceedings of ACL'99: the 37th Annual meeting of the Association for computational Linguistics.

[5] O.Etzioni.The World Wide Web: Quagmire or gold mine.Communications of the ACM, 39(11):65-68,1996.

[6] S.K.Madria, S.SBhowmick, W.K, Ng, and E.P.Lim.Research issues in web data mining. In Proceedings of data ware housing and knowledge Discovery, first International conference, DaWak'99, pages 303-312, 1999.

[7] R.Cooley, B.Mobasher, and J.Srivastava.Web mining: Information and pattern discovery on the World Wide Web. In proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), 1997.

[8] R.kosala, H.Blockeel.Web mining Research: A survey
[9] S.K.Madria, S.S Bhowmick, W.K.Ng, and E.P.Lim.Research issues in web data mining. In proceedings of data warehousing and knowledge Discovery, first International Conference, DaWak'99,pages301-312,1999.
[10] Garofalaski,R.Rastogi,S.Shesadri,K.Shim,Datamining and the web:past,present and future,proceedings WIDM99,Kanas City,USA 1999.
[11] J.Srivastava, R.Cooley, M.Deshpande, P.tan,web Usage Mining: Discovery and applications of usage Patterns from web data,SIGKDD Explorations,Vol.1,No.2,Jan.2000.
[12] J.han and M.kamber, Data Mining Concepts and Techniques, Morgan Kaufmann Publisher, 2001.
[13] R. Agrawal, T. Imielinski, A. Swami. Mining Association Rules between Sets of Items in Large Databases. Proceedings of the 1993 ACM SIGMOD Conference,1993.
[14] Jeffrey Heer and Ed H. Chi. Mining the structure of user activity using cluster stability. In Proceedings of the Workshop on Web Analytics, Second SIAM Conference on Data Mining. ACM Press, 2002.
[15] Eleni Stroulia Nan Niu and Mohammad El-Ramly. Understanding web usage for dynamic web-site adaptation: A case study. In Proceedings of the Fourth International Workshop on Web Site Evolution (WSE'02), pages 53–64. IEEE, 2002.