CHAPTER 2
REVIEW OF LITERATURE

The present research study focuses on quality analytics for automatic evaluation of dynamic web learning environment. It is based on three major dimensions of web analytics, quality evaluation models and metrics identification along with their usage for search optimization. The chapter begins with a brief summary of general approaches towards website development and evaluation. Consequently, the prime leading studies dealing with web analytics and click stream data are reviewed for their importance. The research perspective of quality evaluation models and metrics is discussed in the next section. The chapter concludes with the usage of analytics for optimization of various search techniques to improve the hit ratio.

2.1 Web Analytics Research Perspectives

Assessment always depends on the perspective of the evaluator. Leading researchers and scientists have presented diverse viewpoints to examine the effectiveness of websites. Literature signifies how user-related, function-related, and investor-related views are helpful in evaluation of quality based on the specific metrics.

User-related models concentrate on user-focused factors like website usability and customer satisfaction. Function-related models examine the architectural design and technical quality of websites. Investor-related models are related to the operational performance of websites and evaluate, how well a website is supporting the overall business objectives of a company.

Chiou et al.[13] examine information service approach, marketing approach and hybrid approach. Studies using information service approach work on technical factors such as ease of use, visual design, information quality, and site navigation structure. Many studies with this approach present various heuristics for good site design and user friendly interface.

Marketing-approach focuses on the commercial aspect of websites and evaluates web users as potential customers. It is connected with investor-related view. As information service approach is closely linked with user related and function related
models proposed by Tan et al in [14]. Hence, the combined approach checks the repercussions of both information service and marketing elements in their models.

Palmer in [15] expresses user related views and information service approach towards websites with perspectives of various design elements. The study proposes five crucial website design elements for website usability. The elements are navigation structure, site contents, interactivity, responsiveness, and loading time of a site. Website developers have emphasized these crucial design elements and accordingly have designed appropriate metrics to measure their performance.

Soonawad [16] discusses investor-related views and marketing-approach for its application for the decision-making process of online customers. The study examines the relationship between website components and the customer decision processes. The objective is to gain an understanding of turning site visitors into purchasers.

Irrespective of which approach and viewpoint is applied towards website development and evaluation, site developers need to be objectively measure the effectiveness of chosen site elements. Hence herein lies the importance of web analytics which are to be applied to measure factors to examine parameters through metrics with technical potential.

Web Analytics Association [17] for real time analytics defines web analytics as the “measurement, collection, analysis and reporting of internet data for the purposes of understanding and optimizing web usage”. It takes advantage of the unique measurability of web sites that attempts to gain understanding about website visitors and further uses this knowledge to improve the effectiveness of websites.

Cooper [18] has stated that the field of web analytics was introduced in 1990s, because the public web had begun to expand day by day. In an important research study, Krishnamurthy in [19], reveals that during the past twenty years, web analytics has become a vital part of website development. Amongst the top 500 retail websites, 91% have implemented web analytics to gather relevant information about site users.

Buckin and Sismeiro [20] in their research study have highlighted that in spite of being used by practitioners already for two decades, web analytics is still in the early
growth phase of its life cycle. The field keeps on evolving as more advanced methods are being continuously introduced.

Nakatani et al. [21] emphasize web analytics for understanding online customers, their behaviors, influential design actions and ultimately foster behaviors beneficial to the business to achieve the organization goal. The goal of web analytics is not just to optimize web pages but it is to provide support for the achievement of overall objectives of an organization.

Welling [22] clearly states in his research study that many companies have set up their websites without clear goals but just to fulfill the need to establish online presence, and web analytics can help these companies to fully define the purpose and objectives of their website presence.

Atkinson [23] examines the usage of qualitative and quantitative evaluation methods. According to their research study, qualitative methods include observations, think-aloud and eye-tracking, whereas, quantitative methods include questionnaires, web log data and web analytics tools. This study gives immense coverage in the proposed research for usage of the web analytic tool to measure the desired metrics in quantitative form.

To monitor and improve the quality of web-sites, heuristic expert evaluation and usage testing are the most current laboratory approaches. Elling et al. [24] in their research study discuss the results of these approaches which can be used to revise a website and certain web pages. However, to focus more predominantly on the overall quality of websites, online questionnaires are being used.

In the case of information driven web sites, it seems logical to connect web site quality to usability. The web sites are evaluated based on six criteria of contents, structure & navigation, visual design, functionality, interactivity, and overall experience. Ivory et al. [25] examine contents, interactivity and navigation factors to evaluate website quality in their study.

As per the metrics to evaluate the website quality, three research studies have been intensively studied. These research studies classify site evaluation metrics into three broad categories of functional and navigational issues, contents and style and lastly
contact information, whereas Misisc et al. [26] depend on usefulness, ease of use, and trust factors for evaluating the same.

To evaluate success of websites, Schubert in [27] uses attitude, perceived usefulness, perceived ease of use and lastly comfort level factors. Hong [28] study adds technical aspects to these factors such as efficiency, speed, and reliability which are collectively termed as performance.

Zviran et al. [29] define website quality as user satisfaction and has empirically examined the effect of user-based design and website usability on user satisfaction. They have investigated the relationship among four dimensions for user-satisfaction, usability, user-based design, and website type on commercial websites. The authors argue that results of the metrics could be used for statistical analysis only if the user's intention is not considered.

Clifton [30-31] in his research study examines analytics dealing with gathering visitor website linkage data. There are five main methods which can be used to collect visitor data: Page Tags (client side data collection), Log files (server-side data collection), Hybrid solutions, Network Data Collection devices and packet sniffers. The popularity of page tags has increased as it allows the analysis to be outsourced as a “Hosted” solution.

Page tagging method uses data collected through user browser. The technique is known as client-side data collection. Information is usually captured by Java script code known as beacons or tags, placed on each page of the website. Log file method uses data collected by the web-server, known as server-side collection. This technique captures all requests made to the web-server including pages, images and PDF’s. It is mostly used by standalone software vendors. Hybrid method combines page tagging and log file methods by Clifton [32].

As perceived from this analysis, a disadvantage of page-tagging method is an advantage of log file method. If the two methods are combined they will result in more reliable visitor tracking. Chatterjee et al.[33] discuss Clickstream, Coremetrics, Webtrends, NetTracker and Visual Sciences –HBX analytics using hybrid method.
Google Analytics [34] use page tagging visitor data collection method but it can be configured as a hybrid data collector. Network data collection devices like packet sniffers, gather web traffic information from routers to black box appliances.

The last technique discussed by Plaza et al. [35] uses Web Server Application Programming Interface (API) and Loadable Module. These programs are utilized for extending the capabilities of web servers. The logged fields are enhanced and extended and captured data is streamed to a reporting server in real time web analytic tool.

Arendt [36] explains how, in comparison to other site evaluation tools, usage statistics reported by web analytic tools are more advantageous, because they did not report only what the user tells and what user will do, instead monitors how user actually works with a site. Sen et al. [37] analyze usage statistics provided by web analytic tools to support certain design decisions.

**2.2 Quality Evaluation Models and Metrics**

The current research study is focused on quality analytics. Dr. Tom De Marco [38] has well defined the definition of quality as “Quality is the function of a product that changes the world for the better”. Therefore, it relates to how well a website is designed and how well the design meets the user expectations.

The quality of any website can be measured from programmer perspective and end-users perspective. The programmer perspective focuses on degree of maintenance, security, functionality, whereas the end-users primarily focus on usability, efficiency and creditability. According to researchers Liburne and Devkota [39], website quality may depend on three factors such as task related factors, performance related factors and development related factors.

Nierstrasz [40] discusses that the quality of software is inherently subjective and people can experience different quality even in the same software. According to a previous research, the quality of the website is a true reflection of the quality of software used to develop it. The major focus in all the discussed research work is concentrated over the user satisfaction.
ISO 9126 definition of quality for software products is “Quality is the *totality of features and characteristics of a software product that bear on its ability to satisfy stated or implied needs*”. Website in itself is software with many internal and external attributes within it. It provides an assessment of all these attributes and whenever all these aspects are taken together, it is referred as quality model [41-44].

*Luis et al.* [45] introduce a website quality model for measurement of website quality. Their research work describes the trade-off between the user needs and associated functions to permit the web application for diverse quality contents.

Quality of website is prone to multiple interpretations which are subjective in nature unless they are quantified by a web quality model. A quality model must be identified by a set of measurable attributes which should meet the expectations of the user.

The researchers have explained in their research that the appropriate metrics have to be defined in order to evaluate the quality of website. Therefore, the choice and monitoring of most significant metrics is the most important concern for quality evaluation and analysis. Metrics acts as paramount evidence to success.

During the literature study phase of this research work it has been established that there are number of quality models that exist for quality evaluation of websites. All the models studied are broadly categorized into two categories.

1. ISO Quality Model
2. Customized Models by Researchers

**ISO Quality Model** is an International Organization for Standardization (ISO) framework in cooperation with the International Electro-technical Commission (IEC) and referred as the development of the new standard “ISO 9126 - Information Technology – Software Product Evaluation – Quality characteristics and guidelines” [46].

It defines the quality model which can be applied to any kind of software product and service. In the process of standard revision, two series are established: series ISO 9126 defining quality model and series ISO 14598 describing the quality of the
evaluation process. In several proposals, researchers have used metrics to directly measure quality characteristics from and similar to ISO/IEC 9126. This standard divides quality into six basic characteristics of functionality, usability, efficiency, reliability, maintainability and portability.

A subset of characteristics from the ISO model is a part of the second level study of the proposed model, where each characteristic is broken down into a set of sub-characteristics, which are in turn further differentiated into a set of indicators at the third level. The choice of indicators is based on a set of web quality guidelines [47], W3C standards [48] and the analysis of the existing websites.

Recent research studies reveal that several experts have worked on this basic model to upgrade and customize the quality model for better usage. Researchers have proposed richer quality models through various elaborative ways of decomposing quality characteristics.

Various measurements and evaluations techniques have been explored for different characteristics in order to bring better mechanisms for quality assessment. Many different methodologies have been used for the said purpose [49-53]. Quality frameworks [54], usability guidelines [55], metrics [56-58], criteria [59] are some of the approaches, used for the better model design and usability.

The evolution of many new models is further reviewed by many researchers. Base quality models are appended and updated for generalized web quality assessment at user-centric level for its applicability to diversified end user domains. Similarly, paradigms are also updated for goal oriented frameworks. Such frameworks are further extended for quality models with human computer interactions using internal and external quality factors to reflect the existence and importance of various characteristics. Quality attributes such as reliability, usability, scalability, security and performance are explored for different website domains in order to identify key attributes for new quality models. The attributes are verified on many domains to assure their participation in supporting the conformance of quality in the said domains. Brief discussion on key literature survey conducted to identify the research work to formulate new quality model is explained in Table 2.1.
Table 2.1: Key Literature Survey on Quality Models Upgradations

<table>
<thead>
<tr>
<th>Researchers</th>
<th>Key Area of Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nakwichian and Sunetnanta [60]</td>
<td>A user-centric web quality assessment model for generalized assessment process that can be applied to diverse end-user domains. Quality model is built on ISO/IEC 9126 and IEEE 1061 standards.</td>
</tr>
<tr>
<td>Brajnik [61]</td>
<td>Suggested the adoption of Goal-Question-Metric paradigm as a useful framework to guide the new definition of the quality model.</td>
</tr>
<tr>
<td>Fitzpatrick et al. [62]</td>
<td>Design of quality model with Human Computer Interaction standards for a general set of 12 external and 5 internal quality factors. Defined a set of “enablers” that reflect the existence and importance of the characteristic in question.</td>
</tr>
<tr>
<td>Offut [63]</td>
<td>Systematic analysis for eight quality attributes as reliability, usability, security, availability, scalability, maintainability, performance and time-to-market to identify and discover new model.</td>
</tr>
<tr>
<td>Olsina et al. [64]</td>
<td>Formulated a quality requirement tree containing more than 100 characteristics referring to different website domains of e-commerce, academic sites and provide a descriptive framework to specify these quality characteristics. This requirement tree is rooted on ISO 9126 standard.</td>
</tr>
</tbody>
</table>

Various methodologies and research theories have been used for internal and external evaluation of a website. It has been found through the research study that websites are designed according to its few important technical and functional aspects whereas the quality model is determined by the process of its evaluation, design, implementation and validation involving a variety of methods and tools. Hence, quality metrics need to be defined appropriately in order to carry out these processes.

Quality metrics are the measurement scale, which are useful to evaluate website usage for their understanding and optimization. Web metrics are used to demonstrate the visitor journey once on the website [65-66]. These metrics assess a website for different domains of e-commerce, academic, advertisement and so on. Further, each characteristic is compared against key performance indicators to be used to improve the website and visitor responses.
In this research domain, Lilburne et al. [67] have proposed a Quality Compliance Framework (QCF) consisting of components of quality measurement, quality characteristic and quality sub-characteristic with measurable indicator. These are the terms in percentage value which indicate the degree of an overall quality compliance of the system.

Melody Y. Ivory [68] has proposed the Automated Usability Evaluation (AUE) method. Through this approach, the number of usability terms for evaluating the system are increased. Automated Usability Evaluation has several potential advantages over non-automated methods, including uncovering various types of errors, more consistency, increasing the coverage of evaluated features, enabling comparisons among alternative designs and predicting time as well as error costs across an entire design.

Calero et al. [69] have designed a web classification framework to determine how the classified web metrics can be applied to improve web information access and usage. The web metric classification with 385 metrics using three-dimensional web quality model, including web features dimension, quality characteristics dimension and life-cycle processes dimension have been presented. As a result of this classification, it has been concluded that the triplet usability, operation and presentation with 149 metrics and another triplet usability, maintenance and presentation with 93 metrics are more defined metrics as compared to others. Most of the other triplets, including reuse, have no defined metrics.

Luis Olsina et al. [70] proposed a quantitative evaluation approach to assess the quality of websites called Quality Evaluation Method (QEM). In order to evaluate, compare and rank the quality of websites, authors have applied and worked on a set of activities regarding the proposed methodology.

### 2.3 Approaching Search Engine Optimization Techniques

Evaluation of quality metrics is quantified through web analytics. This statistical measurement through quality models facilitates optimization of techniques for probing the websites and hence leads towards an innovative approach for the same. According to the present literature, there are many techniques available for optimization of search engines such as indexing, preventing crawling, increasing
prominence, white hat and black hat techniques. The proposed research work is concentrated on few of these techniques to improve the key metrics contribution for better quality of web environment at run time. This section is dedicated to the key literature survey in this context.

Zheng et al. [71] associate web analytics and web site performance with the concept of web attention. The researchers propose five dimensional model of web attention in which they state that during the time of information overflow, a successful website needs to effectively catch the attention of the site visitors. This framework is improved in acquiring the web attention in the competitive market. Visibility, popularity, loyalty, depth and stickiness are the five well explored dimensions.

Malaga et al. [72] define Search Engine Optimization (SEO) as an art of designing, developing, modifying and coding web pages to achieve high rankings in the search results and in converting high user traffic. Search engines rapidly change their ranking algorithms with the passage of time but SEO basics remain same throughout. Search engine optimization can be performed by two kind of practices, which are white-hat SEO and black-hat SEO. White-hat SEO also known as ethical SEO includes activities which are carried out according to the guidelines, rules and policies of search engines. Normally, most of SEO practitioners follow it. Whereas, Black-hat SEO refers to the illegal SEO practices against search engine rules and regulations which are derived to get quick search ranking results.

Chung [73] explores search engine indexing processes. The techniques used by these search engines are reading and including website contents and data in corresponding data repositories. Exploring the contents of the web pages for automatic indexing is the key requisite that allows the users including customers and businesses to locate the best sources of their usage. It is pertinent to maintain an index of all the web pages to get maximum ranking through search criteria.

Indexing is the process in which a search engine stores the crawled data through three phases of parsing, indexing document for storage and sorting. The goal of the search is to provide quality search results efficiently. Hence, it uses a query evaluation process for searching [74].
Chapter 2 – Review of Literature

Su et al. [75] suggest SEO as one of the mainstream area on internet. A website can only be optimized and visible when it is well ranked and placed on the first page in search engine results. More the relevance, better will be the ranking. Many research studies in this field reveal that websites, which occupy first five positions on search engine ranking, have maximum traffic and generate more profit.

Other than these, there are several ways and practices to achieve search engine optimization and can be applied in different ways and sequences. But the general practices and guidelines of SEO are same for all kind of website campaigns. For instance, a major search engine like Google has given the guidelines in their help pages to help out webmasters and SEO practitioners. If these general guidelines and SEO practices are not followed then it can lead to failure of SEO campaigns, spam and may be declared illegal for search engines [76].

*Malaga* [77] suggests that it is better to make a brief SEO strategy, work according to the designed sequence in a timely and goal driven manner. It is also important to keep informed about the latest ranking strategy and updates of search engines. Effective SEO strategy requires iterative testing in order to refine it. The author describes four general steps to refine SEO inquiry, which are strategizing, executing, analyzing and optimizing. This motivational research approach has been followed to create a road map of proposed research work for designing research methodology.

*Sen et al.* [78] evaluate quality contents, relevant keywords, strong meta language properties, internal links throughout the site and external links and back links to the site as the basic ingredients for search engine optimization. The main objective of the research is to get the website rank up in the search results of its target keywords. The objectives are achieved by targeting the audience, to check which client wants to sell and promote, which sort of visitors are required to be targeted and in which area, country and time frame it will be completed.

*Tonkin* [79] explains that goal measurement is essential in such type of a research work because it is completely possible to increase traffic to positively affect the business as goal oriented optimization techniques derive the traffic in context to hits, but sometime invariably do not meet the goal as directed for the business.
Wolf [80] explicates how web crawlers play a significant role in SEO. Web page indices are built by web crawler, which are used as a foundation for the implementation of a search engine. Starting from a few seed URLs, the web crawler downloads every page and extracts all hyperlinks found on every visited page and continues by downloading new pages. So it is possible to find every reachable document starting from the seed URLs.

In a similar study by Yamada et al. [81] it is revealed that the pages secured by a password cannot be indexed by a web crawler. These pages are referred to as a deep web, but such pages cannot be found by information retrieval systems, hence it is a big loss of information.

Killoran et al. [82] propose a method to build the search engine through the three-stage processes of crawling, indexing and information retrieval. Search engines find web pages, organize these pages for fast retrieval and serve up the information based on user request. This is an ongoing process because the search engines databases change and are developed constantly. This constant change results in an update on the web pages. The authors, in their research study, have examined these techniques to increase the website visibility.

2.4 Gaps in Literature

In the field of web and quality analytics, as evident from above literature survey, many tools and techniques are in practice to gather web data in real time. Quality models have the potential to quantify metrics for dynamic web learning environment evaluation. The few research gaps which are identified to carry out the proposed research study are as follows:

- Web analytics, although being used to discover and communicate meaningful patterns in recorded data, are not subjugated to value them for optimization.

- Analytical tools, though have potential to record meaningful data patterns and are used for interpretation for predictive analysis, have not been well exploited for quality evaluation in real time.
ISO quality model, provides promising framework for assessment and evaluation of quality of web environment, but these models work in isolation.

Quality models for evaluating website quality and web analytics, both, work in quantifying the values of key performance indicators, but no model exists which shows a strong collaboration between these two approaches.

In most of the existing studies, approaches used to identify the metrics for quality evaluation work in isolation. Combined approach can be exploited to standardize the metrics.

In the research studies, five key ingredients, quality contents, relevant keywords, strong meta language properties, internal and external links have been taken into account for quality evaluation and search engine optimization, but the applied procedures are not focused to monitor them through web analytics and intervene back to the dynamic environment for optimization.

A fixed set of protocols are exploited for search engine optimization. More emphasis is given to the metrics which are directly and indirectly related to static features, whereas, dynamic metrics are not explored completely.

The SEO explores the structure of a website which is subjected to constant change may vary every day. The change of structure due to dynamic behavior of website is analyzed practically.

In view of the above identified gaps, it is inferred that web analytics, quality evaluation models and search engine optimization are the three different domains to be focused on. The potential benefits of applying these three methodologies are better quality evaluation and performance tuning; therefore, the present research is aimed to establish a relationship to find the common key intersectional areas.

2.5 Research Motivation

This intensive literature survey justifies the research potential of this key area and offers sufficient encouragement for pursuing further research. The research motivations are discussed as follows:
Web Analytics, Quality Models and Search Engine Optimization techniques are the key areas in which research work is being undertaken for improving the quality and performance of web applications. They are working in isolation. The current research provides an integrated framework to explore the research potentials of these key areas.

SEO is useful in optimization of the website for better searching, indexing and ranking. In order to achieve it, static features of the website are examined and optimized at design time for better perspective. But, dynamic approach has not been exploited in actual manner. This concern motivates researcher to work for run time environment for quality evaluation and optimization.

Quality models are implemented at the development phase of the website but when the website is hosted, quality is compromised at run time. This concept reveals that quality checks remain active as far as contents, framework and working is in hands, but is beyond control when online audience approach the website through various navigational tools.

Keywords available in the metatag of website are limited and static. The motivation is to make the procedure dynamic so that the website can be optimized at run time by analyzing metatags along with titles, descriptions headings and contents.

Significant relationship between hit ratio and bounce rate motivates towards identifying the key factors due to which website gives high bounce rate. This relationship can be exploited further using exit ratio and search volume analysis in order to identify the strong evidence to restrict bounce rate as well as exit ratio for better performance.

The literature survey reveals that the techniques opted for increasing the visitors on the websites do not consider organic behavior of the visitor. It leads to lower customer satisfaction. User satisfaction motivates to work on the organic behavior of the customer and to work in the direction of increasing genuine visitors rather than increasing the clicks streams.
2.6 Summary

The literature study has revealed that meaningful and high-quality research has been carried out in various acknowledged areas. Many significant approaches such as analytics, quality models and optimization techniques are investigated for quality evaluation and performance tuning but they are working in isolation. In view of the above identified research gaps, the current research aim to provide interoperability among the identified approaches. It provides a refined approach with the potential benefit of development and implementation of quality analytics to automate web learning evaluation. Intensive literature survey regarding analytical procedures, quality models and metrics for quality evaluation and optimization techniques has been carried out. Inferences have been drawn and research gaps in the existing literature have been identified. It has been identified that the research potential of real time web analytics has not been completely explored for optimization purpose in dynamic web environment. Metrics are implemented in quality models but all models works in isolation with each other. Similarly, less work has been carried out in the quality metrics subjected to dynamic changes. These identified major gaps point towards the need to work for a technique which can work in collaboration with respect to quality evaluation of dynamic web environment using real time web analytics for performance enhancement.