Web Analysis and Improvisation Using AI

Vishal Gaikwad[1], Vishal Nikam[2], Sandeep Shukla[3], Bhooshan Shinde[4],

Pradnya Daflapulkar[5]

Student, Dept. of Computer Engineering, MMIT Lohgaon, Pune, India[1234]

Professor, Dept. of Computer Engineering, MMIT Lohgaon, Pune, India[5]

ABSTRACT: Web analysis is a process for measuring web traffic, but it can be used as a tool for business and market research.In this paper, while focusing on website's design styles and critical data of the website is considered, the statistical data is being collected through owner of mentioned website.SEO techniques are used to analyse the grade of the website. After approving the validation of this data, suggestions are provided to the website owner, regarding the flaws of the website. This can increase the fruitfulness of the website.

KEYWORDS: Page rank algorithm; SEO; analyse; suggestions;

I. INTRODUCTION

The world today is emerging as digital arena. Nowadays, people are globally accessible to the internet. From small businesses to large industries, everyone uses internet as a platform to expand their growth. However, many websites contain data and information regarding their respective business or industries but not in correct manner. Either their information is not sufficient or their information is not organized correctly on the webpage. In some of the cases, excess or irrelevant information in the webpage make the ranking of website to downgrade. Another aspect of web analysis is web visualization which also effects the ranking of webpage. Irrelevant tags also lead to the same. Using web analyzing techniques we can organize the information and data in proper and efficient way. Relevant and improved new tags will be suggested usingefficient algorithm. Statistical data like heat map, scroll map, confetti, overlay map for each webpageof website will be generated for better organisation of critical data, and the suggestion will be proposed to placecrucial data on the most viewed part of the website. There are many factors which effect the usefulness and interest of the user on the website. The suggestions and the respective changes made by the owner may enhance the grade of the website, making it pie-in-the-sky.

II. LITERATURE SURVEY

1)The Effect of Implementing SEO Techniques and Websites Design Methods on E-Tourism Development: A Study of Travel Agencies E-Tourism Websites.
Authors:Majid Mohammad Shafiee,Shirin Rahimzadeh,Reyhane Haghighizade
Description:In this paper, with a particular focus on the Web sites that are related to tourism centres and travel agencies, the effect of their websites design styles and implement of SEO techniques on the e-tourism development has investigated. Statistical data have been collected through distributing questionnaire among 70 users of mentioned website and after approving its validation, the results indicate that there is a significant relationshipbetween the effective factors of affecting a proper website design and implement of SEO with the development of tourism.

2)Web sites of the Fortune 500 companies: Facing customers through home pages.
Authors:Chang Liu, Kirk P. Arnett, Louis M. Capella, Robert C. Beatty
Description: The growing popularity of the information superhighway has opened up exciting opportunities for companies looking to, not only maintain their current customerbase, but also to reach new customers. One of the most popular methods to enter into cyber marketing has been to establish a home page orWeb site on the Internet. Almost two-thirds of Fortune 500 companies currently maintain home pages on the Web. An analysis of the content of corporate home pages provides useful insights. Over four-fifths of the companies display products and services (93.2)

and company overview (86.1) information. Roughly three-fourths of the companies present interactive feedback (79.3) and what's new (71.1). Less than one-third (26.2) of Fortune 500 companies provide for online business. An analysis of the data also provides valuable insight into the future trends of home page usage by large business organizations.

3)Applying Web Analysis in Web Page Filtering.
Authors:Michael Chau
Description:Vertical search engines provide Web users with an alternative way to search for information on the Web by providing customize searching in particular domains.However, two issues need to be addressed when developing these search engines that how to locate relevant documents on the Web and how to filter out irrelevant documentsfrom a set of documents collected from the Web. This paper reports the research in addressing the second issue. Traditional approaches, such as a manual approach or a keyword-based approach have their shortcomings. A more promisingapproach is by using text classifiers, but a major problem is that most text classifiers rely on a large number of testing data and do not effectively incorporate Web characteristics into their models. In this research a machine-learning based approach that combines Web content analysis and Web structure analysis is proposed.

4)A Fuzzy Driven Reliability and Relevancy Map for Web Content Search Optimization.
Authors:Ankita Halwasia, Monika Sharma
Description:In this paper, a fuzzy based parameter specific analysis model is presented for web search optimization. This presented work model is applied in three main stages. In first stage, the filtration and the general content mapping is provided. In second stage, the statistical information extraction and its fuzzy based analysis is provided for rank assignment. In final stage, the user interest analysis is applied for improving the ranking results. The implementation result shows that the presented work model has improved the ranking of the pages and more accurate results areobtained.

5)Restructuring Web Search Results by Generating Feedback Session and Clustering Pseudo Documents.
Authors:Bhagyashri Girdhar Salve and R. B. Wagh
Description:The system is used to restructure the web search results by using user search goal and improving the performance of user search goal. In this system the feedback session are used to generate the pseudo documents by using TF-IDF vector. Then K-means clustering algorithm is applied on pseudo documents to generate user search goals after that the restructuring is done by using the user search goals so, user can find the exact information very efficiently and appropriately.

III. PROPOSED SYSTEM

This application aims to analyse the web-page and provide suggestions. Implementation of this application is been done using windows operating system. For handling large amount of data MS SQL is been used. Yii frame work is been used for implementation of SEO.
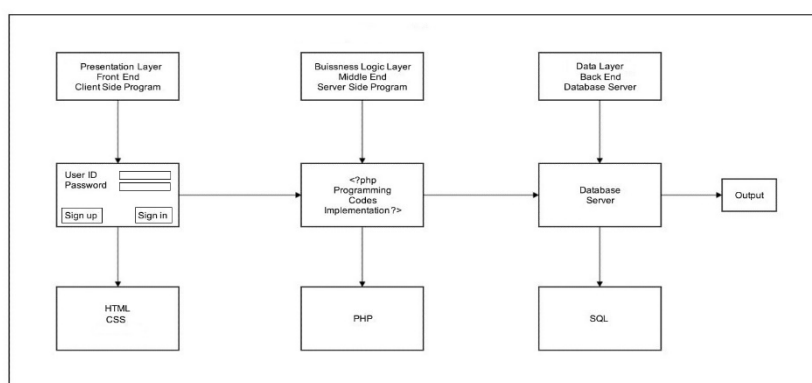

Fig. System Architecture

As shown in Fig. User interface is been provided where user can login and sign up for the application. Data extraction system consist large input of keywords and then we use MS SQLfor storing such large data and simultaneously we analyse the dataset and using SEO technique precised output is given.

IV. MATHEMATICAL MODEL

A= (I, O, f(x), S, F)
where I is set of inputs,
O is set of output,
f(x) is set of functions,
K is system of functions,
S is set of success
and F is set of failure.
I = Q1, Q2…. Qnare n no. of Keywords
K1, K2…Kn are set of formats K
where K1 is Type of keywords,
K2 is on basis of heat-map,
K3 is on basis of scroll-map,
K4 is on basis of overlay,
K5 is on basis of confetti.
Output: Web page ranking and suggestions of keywords.
Success: Increase in web ranking and suggested keywords worked.
Failure: Low increase in rank and suggested keywords didn't worked.

Page Rank Algorithm:

$$PR(pi) = \frac{1-d}{N} + d \sum_{Pj \in M(Pi)} \frac{PR(Pj)}{L(Pj)} \tag{1}$$

where P1, P2… Pn are the pages under consideration, M(Pi)is the set of pages that link to Pi, L(Pj)is the number of outbound links on page Pj, and *N* is the total number of pages.

V.ACKNOWLEDGMENT

VI. CONCLUSION AND FUTURE WORK

As the world is on-line now, every small to large business, firms, education institutes, hospitals etc. use websites for most of their work. Having an efficient and well organised data to grab the users interest is needed. By applying this system, all above things are possible. Hence, by applying effective algorithm and provided suggestions the website proves out to be pie-in-the-sky.

REFERENCES

[1] "The Effect of Implementing SEO Techniques and Websites Design Methods on E-Tourism Development: A Study of Travel Agencies E-Tourism Websites"Majid Mohammad shafiee,Shirin rahimzadeh,Reyhane haghighizade ©May 2016 IEEE.
[2]"Web sites of the Fortune 500 companies: Facing customers through home pages" Chang Liu, Kirk P. Arnett, Louis M. Capella, Robert C. Beatty Information & Management 31 (1997) 335-345.
[3] "Applying Web Analysis in Web Page Filtering" Michael ChauJCDL'04, June 7–11, 2004, Tucson, Arizona, USA.
[4] "Understanding Search-Engine Optimization" Venkat N. Gudivada and Dhana Rao ©Oct 2015 IEEE.

[5] "Analyzing and Classifying User Search Histories for Web Search Engine Optimization" Archana Kurian and Jayasree M2014 3rd International Conference on Eco-friendly Computing and Communication Systems.

[6]Bhagyashri Girdhar Salve and R.B. Wagh, "Restructuring Web Search Results by Generating Feedback Session and Clustering Pseudo Documents" December11-12, 2015, Kurnool, Andhra Pradesh, India.

[7] "A Web-Based Tool for Collaborative Social Media Data Analysis" Xin Chen, Krishna Madhavan 2013 IEEE Third International Conference on Cloud and Green Computing.

[8] "A cloud based web analysis and reporting of vital signs" Yazan A Alqudah; Esam A AlQaralleh 2012 Second International Conference on Digital Information Processing and Communications (ICDIPC).

[9] "Requirement of knowledge centre based on web analysis" Illy Amirah Hassannuddin; Halina Mohamed Dahlan; Ab Razak Che Hussin 2013 International Conference of Information and Communication Technology (ICoICT).

[10] "Application and Research of SEO in the Development of Web 2.0 Site" Chengling Zhao; Jiaojiao Lu; Fengfeng Duan 009 Second International Symposium on Knowledge Acquisition and Modeling.

[11] "How to Improve Your Google Ranking: Myths and Reality" Ao-Jan Su; Y. Charlie Hu; Aleksandar Kuzmanovic; Cheng-Kok Koh 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.

[12] "Research and Analysis of Search Engine Optimization Factors Based on Reverse Engineering" Cen Zhu; Guixing Wu 2011 Third International Conference on Multimedia Information Networking and Security.