A Survey on Search Engine Optimization: Concept and Techniques

Ashwini P. Patil[1], Prof. Prajakta M. Ovhal[2]

[1]*PG Student, Computer Engineering, Dept., Dhole Patil College of Engineering, Pune*
[2]*Assistant Prof., Computer Engineering, Dept., Dhole Patil College of Engineering, Pune*

Abstract- **Basically search engines are answer machines. The search engine returns the relevant or useful results to the person's query on online search. It also ranks the results as per the popularity of the website. SEO is all about optimizing a website for search engines. It enhances website's ranking in search engines. This paper mainly focuses on concept of Search Engine Optimization. It also reviews different SEO techniques adopted.**

*Index Terms-***Crawling, On-Page SEO, Off-Page SEO, Search Engine Optimization.**

I. INTRODUCTION

Search engines typically assume that the more popular a site, page, or document, the more valuable the information it contains must be [19].
SEO is nothing but an optimization of a website for search engines. SEO is a technique for:

i. designing and developing a website to rank well in search engine results.
ii. improving the volume and quality of traffic to a website from search engines.
iii. marketing by understanding how search algorithms work, and what human visitors might search.

SEO is a subpart of search engine marketing. SEO is also known as SEO copyrighting. The reason behind this is the maximum numbers of techniques are used to promote sites in search engines, deal with text [18].

II. WORKING OF SEARCH ENGINE

Search engine performs several tasks to provide search results as per the parson's query. It goes through following activities [18]:

A. Crawling
Process of fetching all the web pages linked to a website. This task is performed by software called a crawler or a spider (or Googlebot, in case of Google).

B. Indexing
Process of creating index for all the fetched web pages and keeping them into a giant database from where it can later be retrieved. The indexing is a process to identify the words and expressions that best describe the page. And then assigning the page to particular keywords.

C. Processing
When a search request comes, the search engine processes it, i.e., it compares the search string in the search request with the indexed pages in the database.

D. Calculating Relevancy
The search engine starts calculating the relevancy of each of the pages in its index to the search string, as more than one page contains the search string.

E. Retrieving Results
Retrieving results is the last step in search engine activities. In this step, results are displayed in the browser.

III. SEARCH ENGINE RANK

When a person searches any keyword using a search engine, it displays thousands of results found in its database. A page ranking is measured by the position of web pages displayed in the search engine results. The web page is having high rank when the web page rank is 1. And this happens when a search engine puts web page on the first position [18].
SEO is the process of designing and developing a website to get a high rank in search engine results. Conceptually, there are two ways of optimization:

i. On-Page SEO - It includes giving proper title to every page, providing good content, putting keywords on correct places, good keywords selection, etc.

ii. Off-Page SEO - It includes link building, increasing link popularity by submitting open directories, search engines, link exchange, etc.

IV. SEO TECHNIQUES

Search engine is a system which gathers information from the internet. Then it organizes and processes information. Finally it displays relevant information to users [1].

Meng Cui [1] in 2011, illustrated the specific explanation of search engine optimization. In this paper [1], the authors introduced the new website building concepts and design concepts to construct search engine optimization.

Shuo Wang et al. [20] in 2011, presented SEO approach based on BP (Back-Propagation) neural networks. The proposed method makes the search engine intelligent and personalized. The method provides optimum retrieval results to the users. The authors adopted algorithm based on BP neural network to optimize personalized search engine. They used the user profiles and the search terms as inputs of the BP neural network to achieve the self-learning. The most suitable search result is provided to user. Such methods provide closer real intention search result as well as avoid the interference. Authors also pointed out about massive research to be done regarding how to collect training samples and how to make numerical of the gathered information accurately and effectively.

Zhou Hui et al. [7] in 2012, discussed the search engine work principle, factors affecting search ranking. They also discussed website search engine optimization method. They also pointed out that website developers need to study the characteristics of own websites and research effective SEO strategy. To enhance the market competitiveness of the website the developers need to constantly update the website content to increase traffic.

John B. Killoran [10] in 2013, focused only on general web search engines. The tutorial delivered lessons that professional communicators can readily implement without specialized technical details. The author mentioned three lessons to find web content and websites through search engines- i] consider the website's competitors and the web content's audiences when analyzing keywords. ii] insert keywords into web text that will appear on search engine result pages. iii] involve their web content and websites with other web content creators.

Archana Kurian [13] in 2014, the proposed method analyzes and classifies user search histories for search engine optimization. In different search engine optimization techniques, the automatically classified query groups are very useful. The proposed method considered a query group as a collection of queries together with the corresponding set of clicked URLs that are related to each other around a general information need. This method proposed a new method of combining word similarity measures along with document similarity measures to form a combined similarity measure. In the proposed method other query relevance measures such as query reformulation and clicked URL concept are also considered.

Sumita Gupta et al. [5] in 2014, presented an outline regarding various page ranking algorithms for academic digital libraries. They also highlighted comparison of these algorithms in context of performance. This comparative analysis encouraged for further required improvement in the related field. As part of future work, an efficient page ranking algorithm in terms of time response, accuracy, and importance of the results; and relevancy of results should be developed and implemented so that the quality of web search results can be improved.

Fawaz Al Zaghol et al. [21] in 2014, analyzed the influence of a local culture and geographical area on search engine ranking. They identified the effect and the relationship of the local society keywords in increasing website ranking. They set a foundation for understanding the search engine optimization in the local geographical area. They suggested the use of social culture keywords to increase the website ranking. Their analysis showed that the SEO is affected by the local geographical area keywords.

Tsung-Fu Lin [15] in 2014, suggested to utilize the technologies of TF-IDF, K-means clustering and indexing quality examination to identify the combination of key words that will benefit search engine optimization. The study demonstrated that it can effectively enhance the website's advancement of

ranking on search engine, increase website's exposure level and click through rate. The proposed method can effectively retrieve combination of key words to benefit the website.

Anchal Kakkar et al. [9] in 2015, explained Google's new calculation HUMMINGBIRD. This paper also focused on the comparison of quality content of different Google search algorithms. The paper interpreted as "SEO can never have a short life compass".

Venkat N. Gudivada et al. [4] in 2015, surveyed current SEO practices and examined possible future directions, to better understand SEO's relationship to business success. Rapidly maturing natural-language translation and machine-learning technologies have fundamental implications for Web search. Each search-engine results page (SERP) presented in response to a user's request contains a series of snippets-clickable links that often include preview text to establish the webpage's relevance to the search.

Rekha Singhal [17] in 2016, proposed and implemented an augmented version of the standard PageRank algorithm by using 'weight' of in-linked web pages. The proposed WIL (Weightage In-Link) PageRank algorithm uses the weights of in-linked webpages to calculate a new score of every individual webpage called WIL-score. Later the webpages can be ranked according to this WIL-score. The proposed algorithm calculates webpage rank score according to importance of webpages based on the extension of standard PageRank algorithm by using the weightage of In-Linked webpages. The proposed WILPR (Weightage In-LinkPageRank) algorithm is targeted to provide higher relevancy in web search results compared to the search results generated by original standard PageRank algorithm.

Santosh Kumar et al. [12] in 2016, contributed two aspects (i) a Dual-Margin Multi-Class Hypersphere Support Vector Machine (DMMH- SVM) classifier approach to automatically classifying web spam by type, (ii) novel cloaking-based spam features which help our classifier model to achieve high precision and recall rate, thereby reducing the false positive rates. The proposed classifier classifies Web pages into four categories, i.e., content spam, link spam, cloaking spam, and combined spam. In future work, they are intended to investigate the aspects of other spam categories for further in-detail categorization.

Dhanashri Ingale [8] in 2016, focused on survey of infer user search goal approaches in prior studies which would help researchers to retrieve efficient results and information. The authors reviewed some of the techniques which are helpful for giving more efficient and relevant information to user search query in less amount of time.

Swati Gupta et al. [6] in 2016, concluded about factors affecting website visibility raising traffic and improving results. This needs to be implemented on websites to find the scope improvement by optimization.

Surbhi Chhabra et al. [2] in 2016, explained a comparative analysis of several SEO techniques has been done based on some inducing factors like N gram indices, Key word & link optimization. The analysis shows that crawling & fetching is the best method because of faster and efficient retrieval of data from search engine.

Joyce Yoseph Lemos [14] in 2017, explained about SEO principles and basic strategies. The paper expresses different techniques that are employed by search engines to improve its results. The paper presents the observation section, which gives the comparative analysis of SEO techniques. The authors analyzed search engine results depend on the various algorithms, which improves web page rank. Web pages are displayed according to their rank, which is calculated by using factor like content, number of incoming and outgoing link etc. This analysis will help to improve the overall performance of search engine in competitive world.

Sylvain Sagot et al. [16] in 2017, proposed a decision support system that allows managers and webmasters to have better visibility on their SEO activity. The objective is to give companies that insource this activity better mastery of the SEO process. They illustrated approach with examples of decision rules created from a typical insourced SEO project scenario. The authors highlighted the role played by the manager and the webmaster in the SEO process. They realized that the outsourced SEO process was transposable to the insourced SEO process, only the human actors had to be replaced.

Rajesh Eswarawaka et al. [3] in 2017, reviewed the factors poignant SEO and also the method a way to do SEO, then defending the approach to bring it into impact by inquiry supported Sis letter of the alphabet. The aim was to assist websites managers to boost

their fames and build a lot of economic effectiveness and social effectiveness. For a close analysis of the strategy and method of SEO, this paper applied Six Sigma management method, to conduct the empirical study of SEO. Six Sigma was initial with success applied in Motorola USA within the late 1980s so extended into several enterprises with an excellent success.

Samedin Krrabaj et al. [11] in 2017, studied and elaborated Google SEO On-page and Off-page techniques. Their analysis targets the site www.studying-in-germany.org, by performing and applying On-page and Off-page optimization SEO factors. To study the overall impact of these factors, measurements are taken using SEERP metrics and tools provided by Google. Finally, we conclude the paper and provide future SEO direction that will improve site ranking in search engines. In 2013, Google introduced an algorithm namely Hummingbird whose main purpose was to understand content context. Two years later, an update of the algorithm called RankBrain part of the overall algorithm for ranking is introduced. This algorithm make use of machine learning techniques for filtering further the results that best fit to the search criteria. The future of SEO will depend mainly in the actual content of the web site and off-page optimization will play a minor role. Thus, building highly qualitative content will play a major role in the future.

V. CONCLUSION

The search engine optimization needs to be implemented or checked with various techniques. The modern techniques- machine learning and deep learning should be also adopted in connection with SEO techniques. This paper explains about the concept of SEO. Later it points out different SEO techniques. It reviews various papers published recently in the field of SEO. The future scope of this work is to survey the SEO techniques purely based on machine learning or deep learning methods and implement appropriate method to get best search results as a part of SEO.

ACKNOWLEDGMENT

REFERENCES

[1] Meng Cui, Songyun Hu- Search Engine Optimization Research for Website Promotion, IEEE, 2011.

[2] Surbhi Chhabra, Ravi Mittal, Darothi Sarkar - Inducing Factors for Search Engine Optimization Techniques: A Comparative Analysis, IEEE, 2016.

[3] Rajesh Eswarawaka, Sai Kiran Kudikala, Sai Charan Kuchi, Viswas Verma- The analysis on Search Engine Optimization supported by Six Sigma Methodology, IEEE International Conference on Innovative Mechanisms for Industry Applications, Page No. 653-658, 2017.

[4] Venkat N. Gudivada, Dhana Rao, Jordan Paris- Understanding Search Engine Optimization, IEEE, Computer Society, vol. 48, issue 10, October 2015.

[5] Sumita Gupta, Neelam Duhan, Poonam Bansal, Jigyasa Sidhu- Page Ranking Algorithms in Online Digital Libraries: A Survey, IEEE, 2014.

[6] Swati Gupta, Nitin Rakesh, Abba Thakral, Dev Kumar Chaoudhary- Search Engine Optimization: Success Factors, IEEE, pp. 17-21, 2016.

[7] Zhou Hui, Qin Shigang, Liu Jinhua, Chen Jianli- Study on Website Search Engine Optimization, IEEE, International Conference on Computer Science and Service System, pp. 930-933, 2012.

[8] Dhanashri Ingale, Dr. M. M. Kshirsagar- A Review on an Approach to Infer User Search Goals for Optimize Result, IEEE Sponsored World Conference on Futuristic Trends in Research and Innovation for Social Welfare (WCFTR'16), 2016.

[9] Aanchal Kakkar, Rana Majumdar, Arvind Kumar- Search Engine Optimization: A Game of Page Ranking, IEEE, pp. 206 – 210, 2015.

[10] John B. Killoran- Tutorial: How to Use Search Engine Optimization Techniques to Increase Website Visibility, IEEE Trans. on Professional Communication, vol. 56, no. 1, pp. 50-66, March 2013.

[11] Samedin Krrabaj, Fesal Baxhaku, Dukagjin Sadrijaj- Investigating Search Engine Optimization Techniques for Effective Ranking: A Case Study of an Educational Site, IEEE 6th Mediterranean Conference On Embedded Computing (MECO), June 2017.

[12] Santosh Kumar, Xiaoying Gao, Ian Welch, Masood Mansoori- A Machine Learning based Web Spam Filtering Approach, IEEE, 30th Int. Conf. on Advanced Information Networking and Applications (AINA), pp. 973-980, May 2016.

[13] Archana Kurian, Jayasree M- Analyzing and Classifying User Search Histories for Web Search Engine Optimization, IEEE, 3rd International Conference on Eco-friendly Computing and Communication Systems, pp. 39-44, 2014.

[14] Joyce Yoseph Lemos, Abhijit R. Joshi- Search Engine Optimization to Enhance User Interaction, IEEE International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), pp. 398-402, 2017.

[15] Tsung-Fu Lin, Yan-Ping Chi- Application of Webpage Optimization for Clustering System on Search Engine –Google Study, IEEE, International Symposium on Computer, Consumer and Control, pp. 698-701, 2014.

[16] Sylvain Sagot, Alain-Jerome Fougeres, Egon Ostrosi- Business Constraints Integration in a Search Engine Optimization Fuzzy Decision Support System, IEEE 19th Conference on Business Informatics, pp. 302-311, 2017.

[17] Rekha Singhal, Saurabh Ranjan Srivastava - Enhancing the Page Ranking for Search Engine Optimization Based on Weightage of In-Linked Web Pages, IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2016), December, 2016.

[18] Tutorial: SEO, Tutorials Point, 2015.

[19] Tutorial: The Beginners Guide To SEO.

[20] Shuo Wang, Kaiying Xu, Yong Zhang, Fei Li - Search Engine Optimization Based on Algorithm of BP Neural Networks, IEEE, Seventh International Conference on Computational Intelligence and Security, pp. 390- 394, 2011.

[21] Fawaz AL Zaghoul, Osama Rababah and Hussam Fakhouri- Website Search Engine Optimization: Geographical and Cultural Point of View, IEEE, UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, pp. 452-455, 2014.