

# Survey Report on Search Engine Optimization

Anindya Jyoti Pal<sup>1</sup>, Unmesha Banerjee<sup>2</sup>, Kanchandeep Ghosh<sup>3</sup>, Hemanta De<sup>4</sup>

<sup>1</sup>Heritage Institute of Technology,  
Chowbaga road, Anandapur, Kolkata 700107, India

<sup>2</sup>Heritage Institute of Technology,  
Chowbaga road, Anandapur, Kolkata 700107, India

<sup>3</sup>Heritage Institute of Technology,  
Chowbaga road, Anandapur, Kolkata 700107, India

<sup>4</sup>Mentor, Heritage Institute of Technology,  
Chowbaga road, Anandapur, Kolkata 700107, India

## Abstract

*Search engine optimization is a strategical technique to take a web document in top search results of a search engine. Online presence of an organisation is not only an easy way to reach among the target users but it may be profitable too if optimization is done keeping in view of the target users as of the reason that most of the time users search out with the keywords of their use (Say; PhD in web technology) rather than searching the organisation name, and if the page link comes in the top positions then the page may be profitable. The work on Optimization Technique describes the tweaks of taking the page on top position in Google by increasing the Page rank which may result in the improved visibility and profitable deal for an organisation. The core methodology used in SEO is to upgrade both content and associated coding of the website to improve its visibility and prominence in organic searches made by the search engines. One of the important tools used in SEO is Data Clustering. Cluster Analysis is a process of grouping the objects, where objects can be physical like a student or can be an abstract such as behaviour of a customer, handwriting. Clustering has been applied to serve various purposes like, to gain insight to data distribution, generate hypotheses, to observe the characteristic and find anomalies, to form the natural classification and even to summarize a data.*

**Keywords:** Search Engine Optimization, PageRank, Data Clustering.

## 1. INTRODUCTION

Search engine optimization (SEO) is the process of affecting the visibility of a website or a web page in a web search engine's unpaid results — often referred to as "natural," "organic," or "earned" results. In general, the earlier or higher ranked on the search results page and more frequently a site appears in the search results list, the more visitors it will receive from the search engine's users, and these visitors can be converted into customers. As an Internet marketing strategy, SEO considers how search engines work, what people search for, the actual search

terms or keywords typed into search engines and which search engines are preferred by their targeted audience. Optimizing a website may involve editing its content, HTML and associated coding to both increase its relevance to specific keywords and to remove barriers to the indexing activities of search engines. Promoting a site to increase the number of backlinks, or inbound links, is another SEO tactic. The working of search engine involves 'spider' which would "crawl" that page, extract links to other pages from it, and return information found on the page to be indexed. The process involves a search engine spider downloading a page and storing it on the search engine's own server, where a second program, known as an indexer, extracts various information about the page, such as the words it contains and where these are located, as well as any weight for specific words, and all links the page contains, which are then placed into a scheduler for crawling at a later date. It was often seen in earlier days that Page ranking were easily manipulated by web master because of shortcoming in search engines algorithm. An evolutionary change came with Google. Google founder Larry Page founded the [Page-Rank](#) algorithm.. PageRank, is a function of the quantity and strength of inbound links. PageRank estimates the likelihood that a given page will be reached by a web user who randomly surfs the web, and follows links from one page to another. In effect, this means that some links are stronger than others, as a higher PageRank page is more likely to be reached by the random surfer.

## 2. SURVEY

Miss. Gayatri Vivekrao Kpase and Dr. V.M.Thakre worked on Search Engine Optimization with Efficient Page Ranking Algorithm<sup>[1]</sup> where they have explored how SEO techniques increase website visibility. For that Web mining technique is used which classify the web pages and internet users by taking into consideration the contents of the page and behaviour of internet user.

Web mining helps the internet user about the web pages to be viewed in future. Web mining is made of three branches i.e. web content mining, web structure mining and web usage mining. WCM is responsible for exploring the proper and relevant information from the contents of web. WSM is used to find out the relation between different web pages by processing the structure of web. WUM is responsible for recording the user profile and user behaviour inside the log file of the web. Their Paper emphasizes that depending on the algorithm used, the ranking algorithm provides a definite score to resultant web pages. A search engine should use web PageRanking techniques based on the specific needs of the users. After going through analysis of algorithms for ranking of web pages compare with the various parameters like methodology, input parameters, relevancy of results and importance of the results, it is concluded that existing techniques have limitations especially in terms of accuracy of results, time response, and importance of the results. An efficient web PageRanking algorithm should overcome these challenges efficiently with global standards of web technology.

In the journal Page Rank based Design and Implementation of Search Engine Optimization<sup>[2]</sup> M Karthikeyan, K Sangeetha have explored the different modern technique that are being used by web masters for better results and they have also covered the influence of SEO on Internet Banking. They have taken up two specific cases one of online SIM and another of Insurance company and they succeeded in creating a much better rank over their period of study. they worked mainly on three propositions. The propositions are as follows:

P1: An increase in search engine ranking will result in a significant increase in the amount of visitors to a web site through the search engine.

P2: An increase in the search engine ranking of a web site will result in an increase in the added value per visitor.

P3: An increase in the number of visitors from search engines to a web site will result in an increase in the total added value of these visitors to the web site.

In the data analysis, all three propositions were tested. In order to draw the conclusions based on reliable information first the figures of Google Analytics have been validated. Therefore the traffic amounts that were measured using Google Analytics were compared with those in Clicky. The results showed corresponding trends in both web analytic tools, which means that it is safe to draw conclusions based on those figures. Based on findings in this research the first proposition was accepted, which means that an increase in search engine ranking actually results in additional web site traffic. Opposite to prior expectations the second proposition was rejected. A higher search engine ranking not only results in more relevant traffic but also attracted more bulk traffic which lowered the average value per visitor. Nevertheless

did the increase of web site traffic outrun the decrease in value per visitor; therefore the overall value that was added to the web site by increasing its search engine rank grew significantly.

In Research of Page ranking algorithm on Search engine using Damping factor<sup>[3]</sup> author Punit R Patel has used data mining techniques for better Search Engine Optimisation in Googles Page Rank The Page Ranking algorithms which are an application of web mining play a vital role to easier navigation for users. In this literature review we have discussed about Web Mining and its categorization, beside this we have explained page rank algorithm and how it employ with

different concept such as number of users that visit the web pages. And also analyze the page rank of web pages for search engine. Based on the survey, we take the different damping factor for analysis and find out that the general scenario of damping factor is 0.85 that means most probably 0.85 values is used for page ranking of web page. In the traditional Google PageRank algorithm, the damping factor is the major element to change the page ranking in hyperlink

diagrams. Analysis results indicate four categories of PageRank based on the damping factor  $d$ . All websites have Minor changes in PageRank value regardless of how  $d$  changes after 0.85 value of pagerank is change slightly.

In High Dimensional Data Clustering Using Cuckoo Search Optimization Algorithm<sup>[4]</sup> authors Priya Vaijyanthi, Xin-She Yang, Natarajan A M and Raja Murugadoss have made an attempt to use Cuckoo Search Optimization (CSO) algorithm to solve the problem of document clustering. The CSO algorithm is experimented with standard benchmark dataset, Classic4 dataset. The quality of solutions generated by CSO algorithm in terms of DB Index was compared with K-means algorithm and Ant Colony Optimization (ACO) algorithm. The results reveal that CSO algorithm is a viable to achieve world class solutions to high dimensional data clustering. Nature inspired meta heuristic approaches are proved to be dominant techniques to attack combinatorial optimization problems in generating near optimal solutions. Cuckoo Search Optimization algorithm is very new in this list and has been proved to be effective optimization algorithm. In this paper, CSO algorithm has been used to find near good quality solutions to document clustering problem. The results were compared with that generated by Ant colony Optimization algorithm and K-means algorithm. It is observed that the performance of CSO algorithm is as good as ACO algorithm in some cases and marginally better in most of the situations. solutions generated by ACO and CSO algorithms were studied.

Ashish Kumar Kushwaha<sup>1</sup>, Nitin Chopde have discussed in Hybrid Approach for Optimizing the Search Engine Result<sup>[5]</sup> have proposed a hybrid model which is the hybrid of Query Recommendation and document clustering, genetic algorithm, model consists of Query Recommendation system in paper learning from historical query logs . This proposed system calculate user's information requirements in a better way by performing query clustering to find the similarities between the two queries, which is based on user query keywords and clicked URLs. After that Generalized Sequential Patterns algorithm is used to generate the frequent sequential pattern of web pages visited by user in each cluster then previously assigned rank score of the web page are modified to re-rank the search result list by using the discovered sequential patterns. The relevancy of the web pages based on its access history is enhanced by rank updation. This can be useful for search engine to optimize the displaying result and able to display the most relevant WebPages with recommendation to user query so user not have to search through list of displayed page and seeking time the of user to retrieve the needed information from the list of pages is reduced by displaying most relevant and use information at the top as per user requirement.

In the Journal Web Information Retrieval Using Genetic Algorithm-Particle Swarm Optimization<sup>[6]</sup> Priya I. Borkar and Leena H. Patil presents a model of hybrid Genetic Algorithm -Particle Swarm Optimization (HGAPSO) for Web Information Retrieval. Here HGAPSO expands the keywords to produce the new keywords that are related to the user search. PSO is an evolutionary computation method, which is clearly different from other evolutionary-type methods that does not use the filtering operation (such as crossover and/or mutation) and the members of the whole population are maintained through the search procedure. In order to find an optimal or near-optimal solution to the problem, PSO updates the current generation of particles (each particle is a candidate solution to the problem) using the information about the best solution obtained by each particle and the entire population. Another component is Information Retrieval System (IRS), that is, a system used to store items of information that need to be processed, searched and retrieved corresponding to a user's query. Most IRSs use keywords to retrieve documents. The systems first extract keywords from documents and then assign weights to the keywords by using different approaches. In this paper, user will search their interest topic through search system. After user enters the keyword, the system will search the term related to that keyword from the database. Then, the result will be presented to the user. From the interface, a user will select interest topic that is most related to the keyword entered before. After that, the keyword will

be arranged in an array to represent the chromosome in binary so that the fitness value for each document can be calculated. Document with high fitness value will be picked in the selection operation.

In Effective Information Retrieval Using Similarity Function: Horng and Yeh Coefficient<sup>[7]</sup> authors Jaswinder Singh, Manoj Chahal have worked on information retrieveal system using the Horng-Yeh Coefficient which are applied using genetic algorithm. In their experiment they put query in search engine and extracted top 10 documents. Encoding documents into chromosome which gives the initial population for input to Genetic algorithm and test Horng and Yeh fitness function with set of parameters: Probability of Crossover (Pc) and Probability of mutation (Pm) to compare the efficiency of Information Retrieval System. In their experiment we also study the effect of different value of crossover and mutation on chromosome. They use crossover probability (Pc = 0.7, 0.8, 0.9) and mutation probability (Pm=0.1, 0.05, 0.01). The new Keyword is now experimentedd and the percentage of improvement is noted. after that they have calculated the Generation and Average Convergence Value at Different Mutation Probability using a second table. All this showed that less mutation rate is best for these queries.

Thus To retrieve relevant information Genetic Algorithm, Information retrieval System and Similarity measure is used. Genetic Algorithm and Horng and Yeh similarity function is used to measure the similarity between query and documents. Horng and Yeh similarity function, vector space model and Genetic Algorithm are applied to increase the efficiency of relevant information retrieval. It is observed that average relevance of documents increases by applying Horng and Yeh formulation in GA. It means Horng and Yeh have refined our search space.

## References

- [1]. Search Engine Optimization with Efficient Page Ranking Algorithm by Miss. Gayatri Vivek Rao Kpase Dr. V.M.Thakre. National Conference on "Advanced Technologies in Computing and Networking"-ATCON-2015 Special Issue of International Journal of Electronics, Communication & Soft Computing Science and Engineering, ISSN: 2277-9477
- [2]. Page Rank based Design and Implementation of Search Engine Optimization by M Karthikeyan, K Sangeetha International Journal of Computer Applications (0975 – 8887) Volume 40– No.4, February 2012
- [3]. Research of Page ranking algorithm on Search engine using Damping Factor by Punit R Patel

INTERNATIONAL JOURNAL OF ADVANCE  
ENGINEERING & RESEARCH DEVELOPMENT-  
2014

- [4]. High Dimensional Data Clustering Using Cuckoo Search Optimization Algorithm by Priya Vaijayanthi, Xin-She Yang, Natarajan A M and Raja Murugadoss (International Journal of Advanced Computer Engineering and Communication Technology (IJACECT))
  
- [5]. IJCSMC, Vol. 3, Issue. 4, April 2014, pg.707 – 710  
Hybrid Approach for Optimizing the Search Engine Result by Ashish Kumar Kushwaha<sup>1</sup>, Nitin Chopde
  
- [6]. Web Information Retrieval Using Genetic Algorithm- Particle Swarm Optimization by Priya I. Borkar and Leena H. Patil International Journal of Future Computer and Communication, Vol. 2, No. 6, December 2013
  
- [7]. Effective Information Retrieval Using Similarity Function: Hornng and Yeh Coefficient by Jaswinder Singh, Manoj Chahal.Vol 3, Iss 8, August 2013  
ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering