Analysing search engine optimization techniques for a full stack web application

G Shivani, Parth Mutreja and Preeti Nagrath
Department of Computer Science, Bharati Vidyapeeth’s College of Engineering, New Delhi
Emails: shivani.gangadharan@gmail.com, pmutreja7@gmail.com, preeti.nagrath@bharatividyapeeth.edu

Abstract. In this paper, we firstly discuss the step by step process to build a fully functioning full stack app and the technologies used while doing so. Following that, a research on various search engine optimization techniques is done for the deployed site of the app. Upon completion of the project and research, we arrived at the algorithm which showed the best results within a given time frame. It was concluded that Google page rank algorithm works the best in this case and also yielded positive results. With that we improved the ranking of our MERN stack application. We also implemented other common techniques for the search engine optimization of a site and have shown results for the same. The factor which significantly improved the ranking of our site was found to be generating more traffic. As the site traffic increased, the ePR score of our site also went up, which was also visible from the increase in the number of clicks.

Keywords: Search engine optimization, SEO algorithms, website development, MERN stack, opportunities, technology

1. Introduction

In this fast growing world of technology, just having an online presence in the form of a website doesn’t suffice. The website should also generate enough traffic, be easily reachable and have a good Google ranking. This will ensure incoming of new users as well as returning of previous users. This is achieved by using search engine optimization, for which there are many methods. With this paper, we present some of those curated search engine optimization algorithms and techniques which the reader can actually implement and get an improved ranking for their website. We will have a web based project, on which we will perform the search engine optimization techniques. The process of building the web based project is discussed along with the work done in improving its ranking.

The aim of the project is to build a website for students in the field of technology, wherein the user enters his/her details and curated opportunities including scholarship programs, internship, community events, etc are fetched. The goal of the research work following that, is to find a fast and easy to implement SEO technique on the deployed site, which most significantly improves the site’s ranking.

Main contributions of the paper:

a. The website building part can serve as a guide to anyone starting to build a full stack application and wanting to learn about its requirements and process.

b. The project description of the full stack MERN application is given with details of each component.

c. Apart from that, we discuss the basic essential SEO techniques and methods which can aid the site get more visibility and can be implemented easily. This work can help even a beginner to understand the most basic SEO techniques that one can perform.

d. We also study and analyse various existing SEO algorithms so this can be helpful when one wants to know which algorithm should be implemented for their use case.
Organization of the paper:

- The paper begins with the abstract where the idea and the aim of our work has been described, followed by the introduction. Here we elaborate on the two major parts of our work i.e., the full stack web application and the search engine optimization of that application. We also briefly discuss all the technologies used in building our project.
- Next, we have the Related Works, wherein we present curated insights from the works of others which helped us gain perspective before we began our work in this domain. It describes the various existing search engine optimization algorithms, the basic and essential SEO techniques, how and when the concept of SEO was developed, etc.
- Following that we have the Methodology section, Here we have explained each and every major step that was involved in building our full stack web application. We then also discuss all the methods we performed for the search engine optimization of our website. These have been represented in the form of flow diagrams for easy interpretation of the reader.
- Then we have the Project Description. This section describes each component present on our website along with its purpose. We have also included screenshots of the same for better understanding and for design visualization.
- Next, we present Results and Discussions. In this section we display all our results, the increase in traffic to our site, the screenshots of improvement in the search results of our site, and more. We also analyse the search engine optimization algorithms and techniques, and understand which one worked the best for us.
- Finally we have the conclusion where we briefly discuss the SEO techniques and algorithms which provided us with desirable results.

1.1 Website

The website is built with the purpose of students in IT, Computer Science or other equivalent fields not missing out on opportunities that can kick start their career. The USP of the website is that it displays curated opportunities for the user based on their education, gender and location. This saves time for the user by giving them only those opportunities which are relevant to them. The type of opportunities on the website range from internships, community events, open source programs, hackathons and mentorship programs. The data is added by the admin, which makes sure there is no spam on the site and that all programs provide top quality exposure and knowledge.

1.2 Search Engine Optimization

The main focus of SEO is to aid websites to organically achieve top placement in search results. It increases the relevancy of the website when the user types the search query on the search engine. Since, the SEO is linked to improve the ranking of the website, the process requires time and skill to implement. During the initial phase, huge investment of time is required for optimisation and maintenance of the website. Depending on the group of keywords used for the optimisation, SEO may take time before the results are achieved. But, the long term benefits of SEO can lead to high traffic of users on the website and better ranking.

1.3 Technologies used

The frontend part of the website has been built on ReactJS, a popular JavaScript library. CSS-in-JS, or in other words, styled-components were also used to easily add functionalities and styling to the components.

The backend of the website has been built using NodeJS and ExpressJS. This connects the website to the database and functions on the server side.
Finally, the database is stored in MongoDB Atlas. This enables us to create multiple rows and fields according to our choice, without creating unwanted fields for each and every row. So altogether, the project makes use of the MERN stack.

2. RELATED WORK

The search engines help to rank the websites on the search result page using various different algorithms. Now such algorithms use some of the key information of the website for calculating the rank of the website.[1] The purpose of the SEO is to use algorithms that aid in internal as well as external optimisation. It takes into consideration on how search algorithms functions and what do people generally search for, in terms of the online customer value and user-centric design layout.[2]

Website design, meta tags, related text, RSS feeds keywords, pages supporting different languages play an important role in internal optimisation and external optimisation. [3] It not only navigates information queries on the internet, but it fills in the gap between the user and the network[4]. Search engine contains five components: Crawler, Query Engine, Indexer, Query Engine and Search interface.[5]

The basic SEO process can be categorised into two types: On-paging and Off-paging optimisation. On-paging optimization is a method used to optimise the website in-depth based on the parameters such as website hierarchy, domain name selection, paging name, website map, robot meta tag. On-paging can be further divided into three parts: Black Hat search engine optimisation (SEO), White Hat SEO [6]. Then Black Hat SEO is a technique to enhance the SERP position of the website, with the help of non-organic methods. This technique is used to quickly enhance the ranking in a search result page. It can be given by cloaking, doorway websites, code swapping and keyword stuffing. White Hat search engine optimisation (SEO) technique uses the collected data including the URL of the website and keywords that define and describe the contents of the website.[7]

In the 1990's, SEO science was begun by the webmasters. The initial study used the web structure to predict the rank, that provided better computing time as well the yield.[8] According to another study, PageRank algorithm, uses the concept of in-links and out-links to rank the webpages. According to this algorithm, if the number of the in-links is more in comparison to other web pages, then the rank of the websites must be more than the other existing pages on the web. The essential pages (having a high rank value) are placed at the higher position in the result list.[9] One of the other methods is Page level keywords method, it includes title, page H1 tag, header, ALT tags and first word of title page, anchor text., Meta description.[10] Google PageRank assigns a rank to each and every recognized website and this data is utilized by search engines to contest the queries of the users.[11] One of the other technique used is Crawler Based Search Engine, it uses a spider to index novel content in the search databases.[12] Web crawler is basically a program which can automatically extract the URLs to suit to the accessor's purpose.[13] After the page is crawled, the contents of the website are indexed. The indexing process refers to cherry picking the words and expression that best describe the page and assigning keyword to each page.[14] There is an essential fault in the PageRank Algorithm as it assigns the complete “importance value” to the webpage based on its quality and quantity. Since, “Page Rank value” is not essentially a query, hence, when a webpage infrequently writes key, the page gets high rank due to the “Page Rank value”.[15] Hill-Top algorithm is a speedy positioning method used to solve the above mentioned problem and this is usually used on the front-end of the PageRank. The request is treated initially by the Hill-Top algorithm. It returns ‘0’ when nothing is searched out. [16]

Hyperlink-Induced Topic Search, also known as Hubs and Authorities, is an iterative algorithm which uses link information to rank web pages. HITS is used to rank pages on a specific topic (the importance of pages depends on the query). HITS divides pages into authority pages and hub pages.[17] The authors of [18] have proposed an algorithm “TrustRank” for the detection of link spam. It is based on an assumption that a good page points to a good page; they seldom point to bad pages (pages containing
spam). Honey Pot are those that provide useful information to users, such as information about latest movies, or information about android programming. [19]. Alexa Rank focuses on the traffic of the website rather than the internal linking within the website. A lower Alexa rank is an indicator of heavy traffic and a better website. The rank allows the webmasters for accurate review of the sites ability to generate income in the long run. [20] RankNet algorithm computes the pairwise score for the respective documents of the pages and makes the use of the weights to enhance the score of the high ranked item or lower the score of low ranked item. The aim of the algorithm is to minimise the inversions in the ranking. [21] WooRank is an algorithm which dynamically grades the algorithm on the scale of 1-100 and represents the internet market efficiency at a given time. It considers parameters such as search engine result pages (SERP) ranking, backlinks, server optimisation and usability for the optimisation process. It has a varying colour criterion: green for cleared (pass), orange for refinement, red for mistake and finally grey color conveys that it won’t hinder the SEO score. [22]

3. METHODOLOGY

We begin by discussing the step by step process for building the web application. It gives the overview of how different technologies and tools were used to fulfill certain requirements during the website building process.

For the SEO part, we begin by discussing the essential and common steps that need to be performed. We also explain how we implemented these techniques for our project. This example can help the reader understand the method better and think how to implement it for their own project or site.

![Workflow of building the website](image-url)

Figure 1. Workflow of building the website
3.1 Building the website
Frontend
ReactJS was used to build the components part of the frontend of the website. The essential components included the navigation bar, dropdowns, information cards and suggestion form among others. Data was passed from the parent to the child component via props, and from child to the parent via call back functions.

Database
Following that, the database was created on MongoDB Atlas and some dummy data was added temporarily.

Backend
Then the server side was built using NodeJS and ExpressJS and was made to run on port 5000. Basically, at this port the data coming from the database is displayed in JSON format.

Connections and state management
Once that is done, the fetched data has to be connected to the client side. This was done using XMLHttpRequest. Then using React Hooks for state management, the data was displayed on the client side website.

Query based fetching was also done with the help of conditional rendering of data in ReactJS itself. Only those opportunities were fetched which matched the user’s entered location, opportunity type, gender and education.

Finally the dummy data in the database was replaced with actual data of various top quality opportunities with all necessary details like name, description, application start/end date, stipend, target gender, etc.

3.2 Performing SEO

![Figure 2. Workflow of performing SEO](image)

Adding keywords in the site
This step is part of the process of building the site. We added certain keywords on different pages of our website. These were the keywords that we expected users to search for, and then land on our site. They included ‘opportunities’, ‘curated’, ‘technical’, ‘open source programs’ among others. An important thing to note here is that these keywords should not be randomly placed in the site’s code but should be a part of meaningful sentences, probably in the About page or Home page of the site.

Adding metadata and title
Following that, we added some meta data to the code. This was done inside the <head> tag of the main index.html file of the website. In a ReactJS project this index.html file would be present inside the public folder. The description of the website was such that it contained a combination of the keywords mentioned above and at the same time, made enough sense to the user and conveyed the purpose of our site in a short sentence. Also, the title of the website plays a crucial role to make the site optimized, so it had to connect to the keywords in some way. So we kept the title as - ‘Opportunist | Curated opportunities in technology’.

Getting the right domain
The very first step after building the first draft of the website, was to purchase the right domain. Now, in order for the website to have maximum visits, it had to have a domain that is easy to remember and at the same time is meaningful to the idea of our website. Since our website is mainly about delivering opportunities in technology we chose the domain name to be Opportunist. It is close to the most used word on our website and also makes sense to the user. The domain suffix was chosen as .tech since all opportunities are in technology and thus it fits very well. The app was deployed on Heroku and the custom domain was bought from Namecheap.

Getting traffic to the website
Among all the steps, we feel that this one helped improve the ranking of our website significantly. We shared the link to our website among many people and the traffic to our site increased. The results of the same are shown in the Results section.

SEO Algorithms
We studied and analysed various SEO algorithms including Google page rank, Trust rank and HITS algorithm. These algorithms and their results have been explained in depth under the Results and Discussions section.

4. PROJECT DESCRIPTION
Figure 3. Homepage of website

The above image shows the home page of the website containing a brief description and the user has two options- to know more or get started. Clicking know more will lead the user to the about section of the page where the purpose is explained further. Clicking on get started will lead the user to the search page, the main feature of this site, shown below.

Figure 4. Search tool

This search tool consists of four dropdowns which take the user’s education, gender, desired opportunity type and location. Based on all these features opportunity details are fetched and shown in the form of cards right below it.

Figure 5. Cards of curated programs

As shown above, the cards contain most of the necessary information that a user might want to know about an opportunity including name, description, target gender, stipend, application start/end date, qualification, location and link to its official website.
Figure 6. Page to suggest opportunities

Shown above is the next feature of the website, where users can suggest an opportunity that can be added to the website. They have to fill this short form requiring most essential details of an opportunity. This data is sent to the admin, who after approving, will add it to the database from where the opportunities are fetched.

Figure 7. Curated blogs on programs

The above image is of the blogs section, as viewed in a tablet or medium sized screen. This section contains links to various blogs written by people who have been previous participants of the respective opportunities. In these blogs, people share their experience, give tips for the application process and other guidance that an aspirant might need.

Figure 8. This blue button shown here is the feedback feature, present on every webpage. Clicking this button will expand it as shown below.
Figure 9. Curated blogs on programs

The user can choose an emoticon based on their experience and also give a brief description if they wish to. This is then sent to the admin.

5. RESULTS AND DISCUSSIONS

5.1 Analysing existing algorithms

5.1.1 Google PageRank Algorithm

It is a link analysis algorithm. It basically assigns a numerical to each and every element of the hyperlinked set of documents. It can be used on any number of entities with reciprocal references. Whenever a page links to another page, it casts a vote for the page. Higher the number of votes that page gets, higher the importance it gets with respect to others[23]. Google PageRank Algorithm is a mathematical algorithm which forms a probability distribution over webpages, so that the sum of the page ranks of the webpages comes out to be 1.[24]

\[ PRank(A) = (1-d) + d \left( \frac{PRank(T_1)}{F(T_1)} + \ldots + \frac{PRank(T_n)}{F(T_n)} \right) \] .... (1)

where,

- \( PRank(T_i) \): Each and every page of the website has a notion of its self-importance. That can be explained as \( PRank(T_1) \) for the first page and \( PRank(T_n) \) for the last page.
- \( F(T_i) \): Each of the pages pervades its vote equally among all of its outgoing links.

For example, the outgoing link assigned to \( C(T_1) \) is one and that for \( C(T_n) \) is \( n \).

- \( PRank(T_0)/F(T_0) \): If suppose a page “x” gets a backlink from page “n”, then the percentage or portion of vote of page “x” is called as \( PRank(T_0)/F(T_0) \)

\( d \)-when the average number of the votes are calculated, to stop the hindrance from the other pages, a multiplying factor “d” is used for dampening down purpose, it is assigned a value of 0.85.

\( (1-d) \)-The \( (1-d) \) bit is used at the starting so that the sum of the page ranks of the webpages comes out to be 1.

The following diagram shows the link structure of our site and how different pages are connected.
5.1.2 TrustRank Algorithm
It is a modification of the PageRank algorithm, but instead of propagating the rank to the web pages, it provides trust to it. It is based on the basic assumption that good pages point to good pages and rarely have connection with the spam pages. Initially, in the process of trust rank, a seed list of pages is made and each of the pages is allocated a trust value in the initial distribution vector \( d \). The value of ‘1’ is assigned if the page is good, else assigned ‘0’. [25]

Function TrustRank

Input
- \( T \) transition matrix
- \( N \) number of pages
- \( L \) limit of oracle invocations
- \( D_b \) decay factor for biased PageRank
- \( PR_b \) number of biased PageRank iterations

Output
- \( t^* \) Trust-Rank score

Begin

// evaluate seed-desirability of pages
1. \( s = \text{SelectSeed}(...) \)

// generate corresponding ordering
2. \( \sigma = \text{Rank}\{1,...,N\}, s \)

// select good seeds
3. \( d = O_N \)
   for \( i = 1 \) to \( L \) do
      if \( O(\sigma(i)) == 1 \) then
         \( d(\sigma(i)) = 1 \)

// normalize static score distribution vector
4. \( d = d/|d| \)

// compute TrustRank scores
5. \( t^* = d \)
   for \( i = 1 \) to \( PR_b \) do
      \[ t^* = D_b \cdot T \cdot t^* + (1 - D_b) \cdot d \]

return \( t^* \)

End
5.1.3 Hyperlink-Induced Topic Search:
It calculates the rank of the website in the offline mode. The Teoma search engine uses the HITS algorithm for the ranking of the webpages. It divides the webpage into two sets categorised as authorities and hubs. Hubs is a set which divides the web pages having outgoing links to other webpages. Whereas, authorities define themselves as having incoming links from other pages. The algorithm recursively generates a hub and authority for each page.[27]

The pseudo code of the algorithm is given below:
Initialize $y(0) = e$, where $e$ is the column vector of all ones.
Below is the iteration that is carried until convergence,
Step 1: $x^{(k)} = L^T y^{(k-1)}$
Step 2: $y^{(k)} = Lx^{(k)}$
Step 3: $k = k + 1$
Step 4: Normalize $x^{(k)}$ and $y^{(k)}$

The equations in step 1 and step 2 can be simplified into:
$x^{(k)} = L^T L x^{(k-1)}$ ....(3)
$y^{(k)} = L L^T y^{(k-1)}$ ....(4)

where,
$L^T L$ is the authority matrix and $L L^T$ is the hub matrix. The above two equations have been used to define the iterative power method for determining the matrix $L^T L$ and $L L^T$

The following table differentiates the above three algorithms on the basis of their description, features, formulae and the results achieved on our website.

**Table 1** Analysing existing SEO algorithms

| Description                  | Google page rank algorithm | Trust Rank Algorithm |
|------------------------------|----------------------------|----------------------|
| Functionality                | A probability distribution over web pages such that sum of the page ranks of the webpages comes out to be 1 | Each page from a list of pages is assigned a trust value in the initial distribution vector $d$. Good page shows 1, else 0 |
| Formula                     | $PR(A) = (1-d) + d(PR(T1)/C(T1) + .... PR(Tn)/C(Tn))$ | $t* = Ab \cdot T \cdot t* + (1 - Ab) \cdot d$ |
| Result on our website       | Improved cPR(rank) because of increased traffic to the site | Provided authenticity to our page and helped improving rank similar to page rank algorithm |

The equations in step 1 and step 2 can be simplified into:
$x^{(k)} = L^T L x^{(k-1)}$ ....(3)
$y^{(k)} = L L^T y^{(k-1)}$ ....(4)
5.2 Search results

The first improvement on the SEO of our website was majorly due to an increase in traffic to our site. As of 15 July 2020, our website did not appear in any search result despite using a variety of keywords, even when we searched the website URL itself on Google. Following that, we added our website to Google domains and shared it among many people thereby increasing the traffic. As a result on 21 July 2020, when we searched ‘opportunist.tech’ on the Google search bar, from incognito mode, our website was the very first result shown. Here’s is the screenshot of results obtained on both the dates:

Figure 11. Search results before performing SEO
Figure 12. Search results after performing SEO

Figure 13. Graph showing the increase number of users visiting the site
6. CONCLUSION
We conclude that the two most important algorithms which provided fast and desirable results were Google page rank and Trust rank algorithm. Other than these algorithms, the SEO techniques that worked for our site were adding a suitable title tag, meta description and appropriate keywords in the website and sharing among people to increase traffic to the site. Directing links on other reliable and high ranked sites to our website also significantly improved our site’s rank and visibility. All these factors put together helped in successfully performing SEO and giving a decent rank to our website on Google.

REFERENCES
[1] Özkan, B., Özcelyan, E. & Kabak, M. (2020) Evaluating the websites of academic departments through SEO criteria: a hesitant fuzzy linguistic MCDM approach. https://doi.org/10.1007/s10462-019-09681-z

[2] McMahon, D. and Griffy-Brown, C. “Developing an effective and affordable search engine marketing strategy for nonprofits”. International Journal of Internet Marketing and Advertising 5, 1-2, 113-130, 2009

[3] Tsuei, H., Tsai, W., & Pan, F. (2020) Improving search engine optimization (SEO) by using hybrid modified MCDM models. Artif Intell Rev 53, 1–16 (2020). https://doi.org/10.1007/s10462-018-9644-0

[4] Ou C (2008) Decryption SEO-search engine optimization and successful strategies of website, vol 8. Electronic Industry Press, China, pp 15–19
[5] Obeidat J., & Hanandeh R. (2013) Constructing Websites toward High Ranking Using Search Engine Optimization SEO, Pre-Publishing Paper

[6] Ghulam A., (2017) On-Page Search Engine Optimization (SEO) Techniques Model: A Use Case Scenario of a Business Entity Website, International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 5 Issue XI, page 3076-3083

[7] Roslina, A.T., & Shahirah, M.N. (2019). Implementing white hat search engine technique in e-business websites. IC4E '19.

[8] Broder, A.Z., Lempel, R., Maghoul, F. (2006) Efficient PageRank approximation via graph aggregation. Inf Retrieval 9, 123–138 https://doi.org/10.1007/s10791-006-7146-1

[9] Sote, A.M., & Pande, S.R. (2014) Application of page ranking algorithm in web mining. IOSR J. Comput. Sci., 47–51

[10] Xing, B., & Lin, Z. (2006). The impact of search engine optimization on online advertising market. ICEC '06.

[11] Higham D.J. (2005) Google PageRank as mean playing time for pinball on the reverse web

[12] W. T. Kritzinger & M. Weideman (2013) Search Engine Optimization and Pay-per-Click Marketing Strategies, Journal of Organizational Computing and Electronic Commerce, 23:3, 273-286, DOI: 10.1080/10919392.2013.808124

[13] Thelwall, M. (2001). A web crawler design for data mining. Journal of Information Science, 27, 319 - 325.

[14] M. J. Kleinberg (1999) “Authoritative Sources in a Hyperlinked Environment,” Journal of the ACM, vol. 46, no. 5, pp. 604–632

[15] Bello, R., & Otobo, F.N. (2018). Conversion of Website Users to Customers-The Black Hat SEO Technique.

[16] Ju Jiehui, “Chinese Search Engines’ PageRank Algorithm and Implementation”. Computer Engineering and Design, 2007,28 (7): 1632-1635.

[17] Kumar, R., & Saini, S. (2011). A Study on SEO Monitoring System Based on Corporate Website Development

[18] Z. Gyöngyi (2004) “Combating Web Spam with TrustRank.”, Proceedings of the 30th International Conference on Very Large Data Bases (VLDB)
[19] Farooq S. (November 2019), "A Survey on Adversarial Information Retrieval on the Web", pp.1-4

[20] Thakur A., Sangal A.L., Bindra H. (July 2011), Quantitative Measurement and Comparison of Effects of Various Search Engine Optimisation Parameters on Alexa Traffic Rank, International Journal of Computer Applications (0975 – 8887), Volume 26– No.5

[21] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender (2005), Learning to rank using gradient descent, in Proceedings of the 22nd international conference on Machine learning, pp. 89–96, 2005.

[22] Shahzad, A., Nawi, N.M., Sutoyo, E., Naeem, A., Naqeeb, S., & Aamir, M. (2018). Search Engine Optimization Techniques for Malaysian University Websites: A Comparative Analysis on Google and Bing Search Engine. International Journal on Advanced Science, Engineering and Information Technology, 8, 1262-1269.

[23] Page, L., Brin, S., Motwani, R., Winograd, T. (1998), The PageRank citation ranking: Bringing order to the web, Stanford Digital Library Technologies Project

[24] Ian Rogers (2002), The Google PageRank Algorithm and How It Works, pp.1-5

[25] Prakash A., & Mustafi D. (June 2013) Fuzzy Logic Approach to Combat Web Spam with TrustRank, International Journal of Innovative Research in Computer and Communication Engineering Vol. 1, Issue 4

[26] Gyongyi, Z., Garcia-Molina, H. and Pedersen, J. (2004), Combating web spam with trustrank, In VLDB Proceedings of the Thirtieth international conference on Very large data bases, VLDB Endowment

[27] Chandranna A.K., (2010), Implementing an Online Version of Hyperlink-Induced Topic Search Algorithm (HITS).