Search Engine Optimization

Prashant Rana[1], Dilip Sah[2], Nived Yogesh Kumar Solanki[3], Mr. Santosh Kumar B.[4]

CSE Department, New Horizon College of Engineering
Bangalore, India
[1]`1nh17cs099.prashant@gmail.com`
[2]`1nh17cs038.dilipsah@gmail.com`
[3]`1nh17cs093.nived@gmail.com`
[4]`santoshkumarcsenhce@gmail.com`

Abstract— **Due to the presence of a massive range of internet sites, the search engine includes a crucial job of providing the relevant pages to the user, Search Engines like Google, use Page Ranking algorithmic program to rank sites in keeping with the standard of their content and their presence over the planet wide internet. programme improvement could be a method of accelerating the probabilities of a webpage to seem within the 1st page of the search result. Since, whenever the buyer searches for info, they supply a specific phrase or a keyword rather than the entire internet address, then the search engine use that keyword to seek out the relevant sites and show it in a very list with the foremost relevant page at the highest. So, a company might use programme improvement techniques to achieve up to its potential client by showing at the highest of the search results. during this paper, we'll be classifying and reviewing totally different technologies for search engine improvement supported their importance and their usage.**

Keywords: Search engine optimisation, Google optimisation, On page optimisation Off page optimisation, Image optimisation, URL structure optimisation

I. INTRODUCTION

The normal search engine works like only how the algorithm feeds up on the server to find specific information and further to display that based on the traffic it gets from the targeted viewers. A search engine basically uses keywords and it's algorithm learns from every letter typed in computer language to turn into what the user is looking for. Taking the next leap forward and introducing only Homepage Links instead of URLs can be a great way to optimize the search engine we use these days. This essentially means that the user (client) of the data rightfully finds the webpage he/she is looking for and doesn't need to navigate to every link displayed on a search hunt to land up at the desired portal.

As search engines are far and away from the foremost user friendly data hunt once it involves any data we have a tendency to are searching for. Here we have a tendency to conjointly introduce videos (Knowing that the amount of users active on Youtube for data searching is double than those on google). In similar words, Youtube is that the next Google. conjointly maintaining the links will be a troublesome task. SEO of knowledge to cloud storage helps such corporations by reducing the prices of storage, maintenance, and personnel.
It may also assure reliable storage of vital information by keeping multiple copies of the data thereby reducing the chance of losing data by hardware failures.

Storing of user data in the search engine feed despite its advantages has many interesting security concerns which need to be extensively investigated for making it a reliable solution to the problem of avoiding local spy of data

II. LITERATURE SURVEY

Since the introduction of the first search engine in the 1990s, the concept of a search engine has grown in importance. The search engine, like email and other common online activities, is considered a basic activity. As a result, search engines are regarded as digital network ecology's concierge. At the moment, modern search engines have excellent access to vast amounts of data. The first prototype of the search engine was created in 1990 by students at McGill University in Montreal, who created a script-oriented content accumulating program that can download multiple files from an FTP directory. Later, the same concept was repeated with a few additional technical details. With the evolution of technology, it is now possible to create massive databases with massive indexes of web sites. Google, Yahoo, MSN, and others are some of the most popular search engines. (Serge Abiteboul and Victor Vianu, 1997)

From the above explanation of the background of search engine optimization, it is clear that search engine optimization has become as common as checking one's email in the daily lives of web users. An experimental investigation by students at McGill University in Montreal started the first prototype of the search engine optimization paradigm. This was later used in World Wide Web services, where web pages are gathered and searched according to requirements. Then, as technology has progressed, large search indexes such as Google, Yahoo, AOL, MSN, ASK, and others have become available. The most popular search engine is Google, as shown in the graph below, and the rest are considerably less popular.

The word SEO is a short form of 'Search Engine Optimization,' according to a study conducted by the Bivings Group on June 18, 2008. In general, the optimization procedure is carried out by abusers in search engines such as Google, Overture of Yahoo, and others, and the web pages of these websites are ranked on the top rankings.

As a result of the foregoing, it can be argued that search engine optimization, or SEO for short, is a way for producing optimal search results for users.

(By Vertexera Inc) Search engine optimization is a method of enhancing a website's visibility in search engines. Selecting a certain axiom or a phrase connected to it boosted this. The search engine optimization deals with data and design concerns that are required to resolve a problem with a site's ranking or rating. The process of search engine optimization is not limited to a single attempt because it necessitates testing using the trace and slip technique, updating on a regular basis, and upgrading the performance level on a periodic basis in order to maintain the site's rank. Organizations usually contract out this duty to companies or individuals that are professionals in this industry for this purpose. It is predicted that the Search engine renders over 500 billion articles. As a result, a company's website may expect to encounter a lot of competition in order to achieve a high ranking.

The requirement for optimization has grown in tandem with the company's effective rate of marketing and client-drawing power, which is made feasible by optimizing the company's websites. (g., Rita Vine)

The following are the advantages of the current developments above traditional search engine optimization: • The process of duplicating the original content of web pages has been minimized, which is a good thing. • increased the crawler's searching capability's speed • Improved efficacy and traffic for the business • High ratings are possible

(Kody Rylas, Dave Young) Search engine optimization has a number of drawbacks and limits. The following are a few of these:
• Competitor restrictions
• Subpages are limited.
• Limitation of crawl ability

• Limitation of duplication
• Linguality

As a result of the preceding investigation of the notion of search engine optimization and its improvements, it can be concluded that search engine optimization has dramatically increased in importance, as it is a very regularly paired habit for netigens. This idea has resulted in a very positive outlook for corporate marketing.

III. PROPOSED SYSTEM

One of the vital considerations that require to be self addressed is to assure the client of the integrity i.e. correctness of his information within the program. because the information is physically not accessible to the user in hand it ought to give the way for the user to examine if the integrity of his information is maintained or is compromised .The system of this search engine is designed in such a way that it displays the Webpage shots as images which when clicked take to the display webpage. Alongside also stores related videos deprived from Youtube to make the user understand better about what they're looking for.

Advantages:

• Avoiding local storage of data, It is important to note that our proof of data integrity protocol.

• By reducing the scattered links on search hunt. • It reduces the chance of being navigated elsewhere as the user can properly identify where he/she wants to go from the webpage shots displayed.

• Not cheating the owner with SEO for the companies own personal gain.

Fig 1. Proposed System

IV. TOOLS USED

A. *Puppeteer*

• Puppeteer could be a Node library that provides a high level API to manage chromium or Chrome over the DevTools Protocol. puppeteer runs headless by default, however may be designed to run full (non- headless) Chrome or Cr. The puppeteer API is ranked and mirrors the browser structure.
• The DevTools Protocol is used by Puppeteer to communicate with the browser.
• Multiple browser contexts can be owned by a single browser instance.
• A browsing session is defined by a BrowserContext instance, which can contain several pages.

• There is at least one frame on the page: the main frame. Other frames may be formed using iframes or frame tags. • The frame's JavaScript is run in at least one execution context, the default execution context. Extensions may be connected with multiple execution contexts in a Frame.
• The Worker has a single execution context, which makes dealing with other WebWorkers easier.
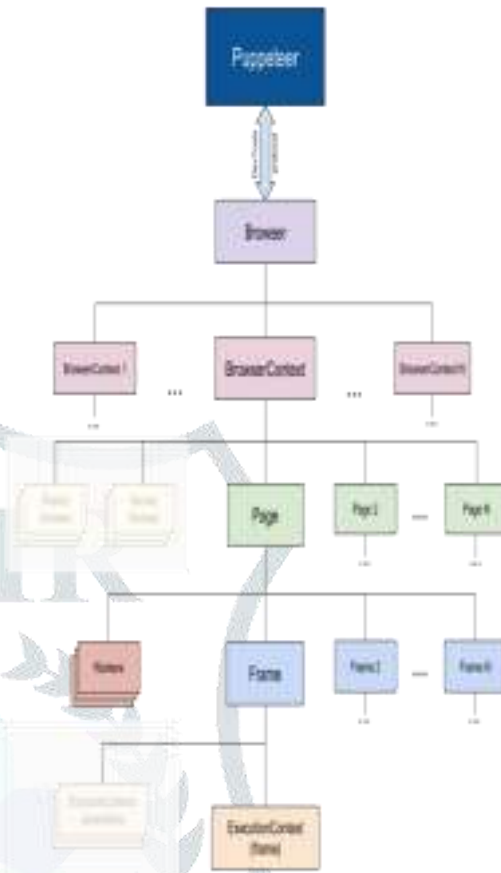

Fig 2.

Puppeteer hierarchical structure

Puppeteer is a browser automation software. When you install it, it downloads a version of Chromium and then uses puppeteer-core to drive it. Puppeteer, as an end-user product, has a number of useful PUPPETEER_* env variables for customizing its behavior. Puppeteer-core is a library that may be used to control anything that uses the DevTools interface. When Puppeteer-core is installed, it does not download Chromium. Puppeteer-core is a library that is entirely controlled through its programmatic interface and ignores all PUPPETEER_* environment variables.

To summarize, the following are the only differences between puppeteer-core and puppeteer:

• When Puppeteer-core is installed, it does not immediately download Chromium.

• All PUPPETEER_* env variables are ignored by Puppeteer core.

When using Puppeteer-core, import the package as:

```
const puppeteer = require('puppeteer-core');
```

The puppeteer module gives a way to dispatch a chromium occasion. The following is a typical example of using Puppeteer to drive automation:

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch();
  const page = await browser.newPage();
  await page.goto('https://www.google.com');
  // other actions...
  await browser.close();
})();
```


Fig 3. Pagination

B. *XAMPP*

XAMPP is an shortened form where X stands for Cross-Platform, A stands for Apache, M stands for MYSQL, and the Ps stand for PHP and Perl, individually. It is an open source bundle of web arrangements that incorporates Apache dissemination for numerous servers and command-line executables in conjunction with modules such as Apache server, MariaDB, PHP, and Perl. XAMPP makes a difference a nearby have or server to test its site and clients by means of computers and laptops before discharging it to the most server.

It could be a stage that outfits a appropriate environment to test and confirm the working of projects based on Apache, Perl, MySQL database, and PHP through the system of the have itself. Among these technologies, Perl could be a programming dialect utilized for web advancement, PHP may be a backend scripting dialect, and MariaDB is the foremost strikingly utilized database created by MySQL. A point by point depiction of these components is given below.

XAMPP is one of the broadly utilized cross-platform web servers, which makes a difference for designers to make and test their programs on a neighborhood webserver. It was created by Apache Companions, and its local source code can be reexamined or adjusted by the gathering of people. It comprises of Apache HTTP Server, MariaDB, and mediator for the diverse programming dialects like PHP and Perl. It is accessible in 11 dialects and bolstered by diverse stages such as the IA-32 bundle of Windows & x64 bundle of macOS and Linux.

C. *Serpstack*

The serpstack API was created to allow for real-time and large-scale scraping of Google SERP data. The simple HTTP GET URL structure takes only a few minutes to set up, and the results are returned in JSON or CSV.

This paper will go over the standards, access, and available API endpoints in great depth. At the bottom of the page, you'll find code samples in a variety of programming languages.

V. WORKING

For the Front End of our system, HTML, CSS and JavaScript has been used, making our system more appealing to our user/client. HTML is used to create the skeleton of the webpage, CSS is used for colouring and styling of our webpage and JavaScript is used to make our webpage more dynamic.

There are two webpages created for the front end. One is homepage which is the base webpage of our system. Here, only a search bar and a submit button is provided as these are the only components required for homepage. A user can enter their search query in search bar and click on submit button to search for results. The other one is Results page which contains our result after we submit our query to the search engine. This page provides us the list of webpages which can also be found on Bing and Google but here it differs on its layout. DadBack.com renders its contents in a Grid View, providing Gallery like feel. Each webpage container contains screenshot of its homepage, title, URL and snippet. Upon clicking on any container, it takes us to our designated webpage.

Pagination is provided at the bottom of the Results page. It's a feature of DadBack.com which allows us to navigate to different pages of our search query. In every page of our search query, we will find different result. Our Pagination is very reliable and can be used anytime we need to navigate to different page of our search result.

At the back end, our system runs a three step process for which it uses JavaScript, PHP and Node js. At the first step, a query is entered in server side website to collect all the website Urls related to our query. Upon clicking on search bar, our server side website searches all the relating website in World Wide Web and stores on the disk.

Then at the second step, all the Urls stored on the disk are retrieved and then our spider i.e. puppeteer navigates to each of those websites to collect their website homepage screenshot. These screenshots are stored on the server disk which will be stored in our database later.

For the final step, the user needs to click on submit which will automatically upload all the related website details along with their corresponding homepage image. Then, our result page shows the success result of uploading the website details. The result page also displays all the homepage screenshots stored in our server.

VI. RESULT AND ANALYSIS

Our process starts from the server side where a search query is provided. On clicking on search button, all the related website results are displayed.
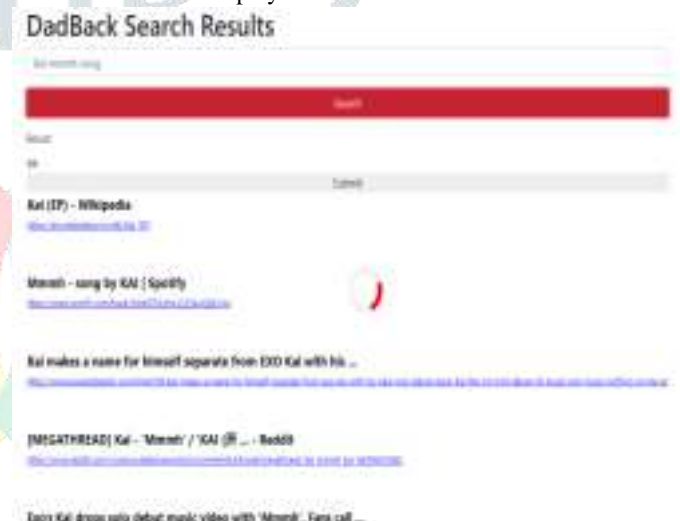

Fig. 4 Server Side search result

The data sets are collected and stored on our disk. The result page shows all the websites' homepage screenshot. These screenshots are stored in a table called images which is linked with our main table websites through primary key and foreign key. The result page displays the success rate and all the data of images table.

Fig. 5 Server side result page

With the success rate of 90%, our system can be moved to client side where the client can use our sytem to search for any query. For the client side two different pages are provided, one is homepage and the other is search result page. The homepage has a simplistic view similar to Google with only logo, input field and a search bar. While the search result page has a unique feature where each search result is displayed as a container containing its homepage screenshot, url and snippet. These containers enlarges as the mouse hovers over it.

Fig. 6 Homepage

Fig. 7 Search Result Page

VII. CONCLUSIONS

In this paper, we briefly describe the motivation for our work and the problem the project is facing. After that, we briefly described our proposed system and how we are going to achieve them. Then, we move onto the tools used in our system. These tools can you used for various purpose to increase the features of our project in the future. Using Serpstack to collect all google results saves us a lot of trouble to collect data. Same way puppeteer can be used to various field like to do data mining and do more of automation tasks. Search Engine has become the basic need of the society to make development. It can be used for a variety of applications. If anyone has any query, their first thought would be to search it in search engine where all the questions are already solved. In the future, the system can be customized to be able to scroll the insides of the container from the search results page only. Upon further utilization of our system, genuine websites will get more attention and in turn this may also increase the likeability of our system. Also, the model can also be improved by showing more dynamic animations.

REFERENCES

[1] A. Gersho and R. M. Gray, Vector quantization and signal compression. Massachusetts, USA: Springer Science & Business Media, 1992.

[2] R. M. Gray, "Vector quantization," IEEE ASSP Magazine, pp. 4–29, April 1984. [3] S. Alkhalaf, O. Alfarraj, and A. M. Hemeida, "Fuzzy-VQ image compression based hybrid PSOGSA optimization algorithm," in IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE, 2015, pp. 1–6.

[3] Serpstack api Documentation [Online] Available: https://serpstack.com/documentation

[4] SEO/Search Engine Optimization and marketing Available:https://searchengineland.com/guide/what-is-seo [5] C.-C. Chang, T. S. Nguyen, and C.-C. Lin, "A reversible compression code hiding using SOC and SMVQ indices," Information Sciences, vol. 300, pp. 85–99, 2015. **[6]** "Particle swarm optimization," in Proceedings of the IEEE International Conference on Neural Network, vol. 4, pp. 1942–1948, 1995.

[7] X.-S. Yang, Firefly algorithm, Nature-Inspired Metaheuristic Algorithms. Luniver Press, 2008. [8] Puppeteer Documentation and how to use it [online] Available: https://devdocs.io/puppeteer/index#class browsercontext

[9] Learning SEO and building SEO friendly site[online]. Available: https://moz.com/learn/seo/what-is-seo [10]SEO starter guide [online]. Available: https://developers.google.com/search/docs/beginner/seo starter-guide