

A Search Engine Basic Optimization Review for Programmers

R. R. Bhawalkar¹, Shalini Kumari², Saurabh vyas³ and khushi Pandey⁴

¹YCCE Nagpur /Information Technology, Nagpur, India

Email: bhawalkrarita1991@gmail.com

²⁻⁴GHRIET/ CYBER SECURITY,

Email: {shalini.kumari@ghru.edu.in, saurabh.vyas@ghru.edu.in, khushipandey1133@gmail.com}

Abstract— Web search engines have made life easier for humans. However, since the web expands daily, it is getting harder to locate the appropriate information about user search queries. A search engine looks through a vast amount of web pages in its database to get relevant results for queries connected to information searches made by users. When developing software, developers, researchers, and students spend a lot of time using popular search engines like Google to locate pertinent materials. In order to shorten the time needed to locate pertinent materials on the Internet, this paper outlines the creation of a search engine specifically for programmers. webpages with guides, illustrations, solutions to problems, etc. Using keywords that are available in several programming languages, different programming languages are crawled and indexed. Number of examples/solutions and lines of code per example are used to rank web sites linked to user queries. The efficiency of the suggested strategy on general search engines is supported by experimental data. Not every user has access to every online page. A search engine (SE) that returns results pertinent to the user's search query is therefore necessary.

To achieve this, a method known as Search Engine Optimization (SEO) is applied. Search engine optimization (SEO) is a technique that makes use of search engine principles to give more relevant webpages higher rankings. A variety of page ranking algorithms can be used to achieve this. This essay reviews SEO strategies that help give consumers of the Internet useful information.

Index Terms— Keyword analysis, meta tags, off-page optimization, programming languages, web pages, and generic search engines.

I. INTRODUCTION

One tool for searching the World Wide Web for information is a search engine. Search engines are used for a variety of tasks, including finding material (text, audio, video, image, etc.) and making reservations and ticket bookings. A search engine operates by utilizing a collection of keywords, or search queries. SEO is a procedure that increases the likelihood that a website will be accessed by using search engine optimization (SE) principles to obtain higher ranks for web pages or websites. With the advent of Google, Yahoo, Bing, and Inquire about SEO innovations and principles the course of growth. In order to systematically present search results to people, SE is a strategy that relies on a number of approaches utilized by computer programs for gathering, organizing, and analyzing data from Internet search information. The four functional domains of a SE are user, indexer, hunter, and searcher. Not only SE Users' convenience is a website's primary purpose, although it is also one. A useful tool for studying how people utilize the internet. Users can locate Target information with speed and accuracy thanks

to effective SE. Major Internet search engines like Google, Yahoo, and Bing are available right now. To obtain information about a website, search engines trawl through its content. Consequently, search engine optimization is required for websites.

Website will rank higher in search results as a result of this. The terms "crawlers" and "spiders" refer to SE programs. Web content and pages are retrieved using a crawler. Web scavenger starts looking at a "seed" list of URLs. The crawler visits URLs, breaking each one apart as it goes. one of the page's hyperlinks and adds them to a crawled URL visit list Search engines scour the internet to identify boundaries. Spiders: Numerous webpages that include links to other webpages elsewhere can be accessed by spiders. It shuts when a link to an external page is viewed. A spider pays a visit. Make a website search index and examine the pages and additional content it contains.

It analyzes data using a variety of algorithms before returning findings associated with the keywords.

The relevancy of the results is determined by additional ranking techniques. Search engines are generally split into four groups viz. search engines that use crawlers, search engines that use directories, hybrid search engines, and meta search engines. Search engines with a crawler-based architecture arrange the content of websites in a logical order. Programs referred to as spiders or bots are used to crawl websites. The spider locates the needed webpage, downloads it, examines its content, and saves the information in the database. The engine looks up 13102.3456789410.1234560.00.0.00000000000000.00.0.000000.0.0.0.01234567890.

000000000000.0000..00123456789 terms in the database when a user conducts a search and displays a list of results. Human editors are responsible for verifying and approving submitted webpages for directory-based search engines.

Site categories are determined by human editors, after which directories within that category are entered into the database. Editors review websites and provide rankings to them according to the content they contain. Crawlers and directory-based search engines are combined to create hybrid search engines. Users have the option of searching the web or the directory.

When displaying results on the Internet, a meta search engine pulls information from other search engines. In order to generate its results, it simultaneously receives input from users and sends search requests to other search engines. After being gathered, the data is prepared and shown to the user.

We provide a search engine concept for programmers in this paper. Using this method, we are taking information out of the webpage according to its attributes. These characteristics include the amount of code on the webpage, the quantity of examples, the presence of a virtual IDE, tests, and quizzes, among others.

Owing to the large number of websites, search engines like Google play a vital role in directing users to the relevant pages. They do this by using algorithms called page ranking algorithms, which rate web pages ba00...123212122200.123456879845612230..0.0.00.1346970*97025130130.0346977*sed on the caliber of their content and their visibility on the internet. Increasing the likelihood that a webpage will show up on the first page of search results is known as search engine optimization.

When a customer uses a specific phrase or keyword rather than the whole web URL to search for information, the search engine uses that keyword to discover related web pages and displays them in a list, with the most relevant page appearing at the top. Therefore, by ranking well in search results, a business can reach out to potential customers by utilizing SEO strategies. Based on their significance and use, several technologies for search engine optimization will be categorized and reviewed in this study.

II. LITERATURE REVIEW

The study put forth a novel method that improves the efficiency of user search data. The relationship between searches, documents, and user inquiries is provided by this mechanism. Take user queries and semantic document structure into account as well. The suggested method produces better results than the earlier methods. An enhanced algorithm for page ranking was introduced in the study. Page rank is now determined by the new algorithm by counting the number of links that visit a page. It has an improved PR algorithm called VOL that produces better outcomes than the previous one. A page with more visits from incoming links has a higher rank value than pages with fewer visits, according to the results, which also demonstrate that VOL performs better than the original PR algorithm. Provide a mechanism for calculating the quantity of link visits to web pages and comparing the VOL and PR algorithms. The study outlines a novel algorithm for determining web page rank based on various factors. The suggested algorithm is known as the HITS algorithm or modified HITS. It is created by expanding the HITS algorithm's capabilities. Web page rank is determined by taking into account six elements.

The Intelligent Search Method (ISM), a novel approach, is presented in this study. It created a brand-new method for indexing webpages with ISM that involves interpreting search requests before indexing the sites in accordance

with the interpretation. Also covered were the constraints and current techniques, as well as the different link analysis algorithms, such as PR and WPR, HITS and CLEVER algorithm. A Content Based Hidden Web Ranking Algorithm (CHWRA) was proposed in the paper. The suggested algorithm possesses four unique characteristics. This approach attempts to address every factor that influences a web page's popularity, whether directly or indirectly.

An ordered concealed web search result set is produced using this technique. The CHWRA algorithm produced the intended outcome. The phrase "document search engine using term similarity vector space model" describes a technique that uses search query semantics rather than keywords directly. He developed the Term Similarity Vector Space Model (TSVSM) in order to address this. This weighted model examines each word's relationship to other terms in the document in addition to comparing words in the input query. The cosine angle is used to check for vector similarity, and the precision and recall values are used to check for accuracy. Every document is read by the system in PDF format. It is assumed that every document has many representations, one of which is the quantity of terms contained within the document. The approach that is suggested in "A Code-Centered Cluster-Based Approach to Finding Online Support Forums for Programmers" examines postings on internet forums that have file attachments that contain data. Every post made on the online forum is duplicated by the system and saved locally. Every post is clustered according to its similarity to other posts and indexed using its TF-IDF score. In order to make functions and methods understandable to all readers, "Vector Space Search Engine for Web Services" suggests a methodology for programmers to name their functions and methods based on their functionality.

By using this method, a search engine is created that compiles the information and determines which match best suits a given request. Because WSDL (Web Service Definition Language) files are XML files that provide all the information required to describe the web service, this solution uses them as the source for the data repository. Crawlers, file sharing, and T-model cross-referencing are used to extract WSDL files. This method makes use of a vector space engine, also known as a term space, in which every WSDL file is represented and every keyword denotes a new dimension. Each document or WSDL file is rated into a vector by computing the cosine value of the frequency of each term using TF-IDF.

In "Towards an Intelligent Code Search Engine," a code search engine that combines document browsing with code example discovery is proposed.

The four main modules of the established system are summary, representation, diversity, and ranking. A collection of code samples is established using an already-existing code search engine, and the examples are then condensed into helpful snippets. Subsequently, each compressed code instance's semantic properties are extracted for ranking and clustering. To give developers a variety of usage patterns, clustering is done. The most representative code examples are then ranked. In order to guarantee programmer efficiency, they incorporate summaries of excellent code examples from the web into API documentation. "Design and Implementation of a Topic-Oriented Search Engine Based on a Multi-Agent System" emphasizes the significance of enhancing general search engine performance and conducting more precise online searches for commercial and financial information. The architecture and operation of search engines.

III. ARCHITECTURE OF SEO

For many businesses, the internet improves the business platform; it modifies work plans and tactics and yields better results. In order to attain their objectives and generate revenue, modern marketers must possess a thorough understanding of SEO. SEO Architecture is the practice of monitoring and motivating webmasters, both internal and external, to enhance their websites. SEO needs to be described holistically, which means that your website (as a whole) must be considered while defining, characterizing, and explaining SEO. Various characters depicted in the figure have distinct functions within SEO. By utilizing them, we can increase profit and business performance.

Categories of search engine optimization

A. Optimizing Off-Page Images

B. Optimization of the Page

Utilizing SEO strategies, websites can increase their traffic by ranking well in search engine results pages (SERPs). Numerous SEO strategies can be used to raise a website's ranking in the search engine results pages.

A. Seo Off-Page:

This will increase the visibility and popularity of your website on the internet. It also aids in raising your website's SERP ranking. One type of off-page SEO is using techniques to get more traffic to your website that you may

disable. Included in off-page optimization are the following methods. *Directory Submission*: The links to online directories and internet directories on this page are submissions to directories. The website's URL, title, description, and other data are submitted during the directory submission process.

Numerous directories, some free and others paid, are accessible; the directory submission includes a little additional information about the business, its goods and services, and its website. The greatest method for keeping bookmarks for internet sites organized and stored is through social bookmarking. In order to retrieve links from their own site, SEO players search for pertinent social bookmarking sites to bookmark.

Instead of learning the web page address by heart when we come across an intriguing website, we may essentially keep the address as a "bookmark" in our browser. We are saving to a website that we can visit from any computer in the world, much like when we save interesting links to our browser. However, the 'social' component is what makes social bookmarking what it is. They can all show up on other bookmarks. This indicates that the content we seek out has previously been bookmarked by users. Forum posting is a link-building strategy where quality forum postings are linked to websites.

Using relevant specialist discussion sites, an attempt is made to re-connect the examination in this strategy. Websites and online pages can benefit from the traffic generated by web forums. As a result, while responding in online forums, SEO experts sign their responses with a link to the website.

In order to obtain high-quality traffic and backlinks to our website and web pages, posting in forums is a crucial and effective SEO link-building strategy. *Submission of Press Releases (PRS)*: This involves posting the most recent news on the relevant websites. This is how we deliver news about what's going on in our company. Backlinks from news websites, a favorable ranking in Google News, and extensive media coverage are just a few of PRS's many benefits. We make sure that our press release submissions are interesting to read when we draft them. You should try writing a press release similar to a newspaper story.

Google indexes and ranks PRS websites highly. A press release we submit may be ranked in as little as one week. A well-written PRS can improve our website's search engine rankings and raise the likelihood that we will receive links from reliable websites. A novel product would be a suitable item for our press release submission.

A well-written press release will draw attention to the product's advantages. An excellent press release can boost our website traffic, open up new business opportunities, and generally make our company better.

Classified Submission: Online classified ads are uploaded to websites that are connected to classified ads in this submission. It's also the most effective way to obtain backlinks. Classified ads on websites pique the interest of the intended visitor in a tiny but effective way. Online classifieds, however, aren't the same as traditional advertisements or business model advertisements. Furthermore, it aids us in search engine optimization by increasing traffic to our highlighted website. Submissions for classified ads boost website traffic.

B. On-Page Seo :

"On-Site SEO" is another name for On-Page SEO. It is the process of making numerous website optimizations that alter the website itself and have an impact on search engine ranking. keywords, URL structure, title page, header, etc. These are included in on-page SEO. Your website's information is provided by on-page SEO. That method is a part of it.



Fig.1

A. *Term Location*: The search engine crawler looks for the term in anchor text, URLs, <title>, <header>, <Alt>, <meta>, and <body> tags, among other tags.

B. Title Tag: The most important ranking criteria is the title. The title tag of a website is typically used by search engines as the primary component of the page that lists the website. The frequency of keywords on a web page divided by the total number of words on the page is known as keyword density. One important optimization consideration should be the frequency of keywords in the body tag and title tag. For a website to rank better, the keyword density should be between 2% and 8%.

C. Keywords in URLs: Adding keywords to this page will help web crawlers find it more quickly. Different domain name suffixes,

such as gov or edu, are preferred by web crawlers. In the same way, short URLs are better for SEO.

D. Keywords in Meta Tags: Meta tags include helpful page or website descriptions that accurately represent the content of web pages. If the meta description tag contains relevant keywords, the website will be crawled.

E. Keyword in Alt Text: Indicates alternate text for text or images contained in the Alt tag. A text description linked to an alt tag that provides the image with the same fundamental information and justifies it. The main text that the image conveys is reflected in the brief yet informative alt tags.

F. Search terms in Anchor Text: Indicates the page to which the anchor text has a link. Link text on pages that are based on search terms is frequently more highly ranked. The website's title ought to accurately convey the content of the page. By employing the title name, the user quickly and easily understands the website's concise content. An appropriate headline length yields positive outcomes.

A URL, or uniform resource locator, indicates where a web page is located on the internet. To help the crawler locate it efficiently, the URL contains search terms. SE idealizes URLs that are shorter.

IV. EXISTING TECHNIQUES

Google uses the Page Rank (PR) Algorithm, which was developed by Larry Page and Sergey Brin. The link structure of the web affects this method. The page is divided equally among its outgoing links by this page rank. The Weighted PageRank Algorithm is a better algorithm than the PR method. Although it uses links as well, its algorithm does not treat page rank equally. Popular pages are assigned a higher page rank. Page rank is assigned based on inbound and outbound links to pages. The Hyperlink-Induced Topic Search (HITS) algorithm is known by this acronym. John Kleinberg is the source of this algorithm. Another name for this technique is the link analysis algorithm. Hub and authority are the main factors considered while ranking web pages. The HITS algorithm increases the number of web pages that the search engine returns in response to a user's query when the user submits one.

The query determines the hit algorithm to use. The HITS style algorithm is used by the well-known Twitter website. Hubs: Pages pointing to authority are represented by hubs. Authority: Authority is a reliable source of knowledge. Web Crawling Agent: This type of program is in charge of sifting through webpages to retrieve data. This method makes use of current general-purpose search engines to produce a whole repository of potential websites. Search results are manually combed through to identify websites including tutorials, virtual IDEs, code snippets, and examples of code.

Typically, these website parts are contained under distinct CSS classes or IDs that are used consistently throughout the website. Therefore, a website's code samples, editable examples, tutorials, etc. will all go under one CSS class. The website's URL is kept in the URL repository, along with the related reference links to the CSS and X-Path selectors. Multiple crawlers, crawler coordinators, URL parsers, and URL database components make up a web crawling agent. A manually selected collection of hyperlinks to different parts of a webpage and URLs of programming websites are kept in a URL database. The crawling operation is initiated by using the URL as a seed page. In order to prevent redundant data collection, crawler coordinators collaborate with several crawlers in real time.

Each page is iterated through by the crawler, which gathers information on the overall amount of lines of code, examples, number of editable tutorials, and page headers for each webpage on the programming related website.

Information Extraction Agent: Information Extraction Agent consists of TF-IDF Calculation, TF-IDF Database, Pages Extraction, and Information Database components. Data extraction and suitable attribute separation are the responsibilities of an information extraction agent. *User Query Processing Agent:* A user's query will be tokenized and sanitized of any special characters when it is submitted to a query processing agent. The same TF-IDF functions are used to process queries and generate vectors with the same dimensions as the code sample vectors. Examples of ranked codes: The technique used in the paper makes websites with more noteworthy instances appear higher in search engine results. The goal of this ranking is to align with the cosine similarity values that the user process agent provided.

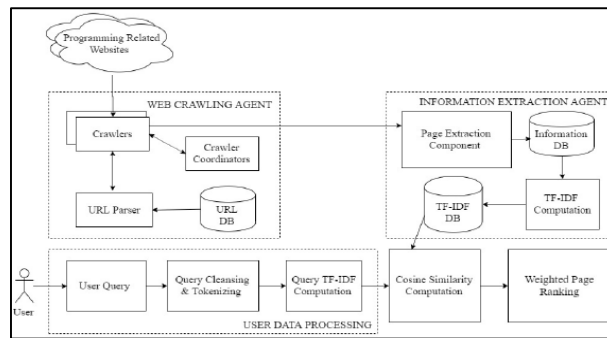


Fig.2 Proposed System Architecture

The information extraction agent will collect many features, including the quantity of editable instances, instances on the page, lines of code, and executable instances, which will be used in the ranking process. Programmers can be sure that they will locate the necessary code examples in the top-ranked examples on pages that provide both editable and highly-ranked examples. The correctness of the examples they give is indicated by the lines of code. Web pages with a lot of search results, editable examples, and more lines of code that match the determined cosine similarity values are given greater weight throughout the ranking process.

V. SEO TOOLS

One tool for suggesting keywords is Google AdWords. A list of comparable terms to the supplied keyword will be produced by this tool. Based on the frequency with which a keyword appears in Google searches, the resulting list is sorted by keyword popularity. Using this tool might help you determine which terms are best for advertising or search engine optimization. You can get the tool at <https://adwords.google.com/select/KeywordToolExternal> for free.

Google Analytics (GA) is an enterprise-class online monitoring technology that provides us with detailed information about the activity on our site and the efficacy of our advertising. We can examine and evaluate our traffic thanks to strong, adaptable, and user-friendly features. We track average time on site, traffic sources, and other information with Google Analytics.

VI. CONCLUSION

Methods of on-page and off-page optimization are crucial for raising a website's ranking. Every optimization technique has benefits and uses of its own. This study reviews the several approaches that can be used to make specific webpages or entire websites search engine friendly. We plan to continue this research by creating an accurate and efficient SEO system that will help websites rank higher in search results. The Programmer's Search Engine helps researchers, developers, and students locate pertinent programming materials online by showcasing the efficacy of a tried-and-true methodology along with the ideas of text processing techniques, crawler-based search engines, and the stability of vector space models. via the outcomes of the trial. The suggested approach can significantly cut down on the amount of time programmers spend online and contribute to more development.

REFERENCES

- [1] R.S. Bhandari and A. Bansal, "Impact of Search Engine Optimization as a Marketing Tool", Jindal Journal of Business Research, vol. 7, no. 1, pp. 23-36, 2018.
- [2] K. Kanathay, R. S. Thakur and S. Jaloree, "Ranking of Web Pages Using Aggregation of Page Rank and HITS Algorithm", International Journal of Advanced Studies in Computers Science and Engineering, vol. 7, no. 2, pp. 17-22, 2018.
- [3] M. N. A. Khan and A. Mahmood, "A distinctive approach to obtain higher page rank through search engine optimization", Sādhanā, vol. 43, no. 3, pp. 43, 2018.
- [4] D. Dandwani and M. N. Punjani, Search Engine Optimization (Working of Page Ranking), pp. 2040-2046, April 2016, [online] Available: <https://www.irjet.net/archives/V3/i4/IRJET-V3I4402.pdf>.
- [5] M.R. Baye, B. D. L. Santos and M. R. Wildenbeest, "Search engine optimization: what drives organic traffic to retail sites?", Journal of Economics & Management Strategy, vol. 25, no. 1, pp. 6-31, 2016.
- [6] S. Gupta, N. Agrawal and S. Gupta, "A Review on Search Engine Optimization: Basics", International Journal of Hybrid Information Technology, vol. 9, no. 5, pp. 381-390M, 2016.

- [7] L. Kumar and N. Kumar, "SEO Technique for a website and its effectiveness in the context of Google Search Engine", *International Journal of Computer Science and Engineering (IJCSE)* Vol. 2, pp. 113-118, 2014.
- [8] P. Rani and S. Singh, "An Offline SEO (Search Engine Optimization) Based Algorithm to Calculate Web Page Rank According to Different Parameters", *International Journal of Computers & Technology*, vol. 9, no. 1, pp. 926-931, 2013.
- [9] A. Jain, "The Role and Importance of Search Engine and Search Engine Optimization", *International Journal of Emerging Trends & Technology in computer science*, vol. 2, no. 3, pp. 99-102, 2013.
- [10] E.D. Ochoa, "Analysis of the application of selected search engine optimization (SEO) techniques and their effectiveness on Google's search ranking algorithm", Diss, pp. 34-40, 2012.
- [11] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine", *Computer networks and ISDN systems*, vol. 30, no. 1-7, pp. 107-117, 1998.
- [12] What are Different Types of Search Engines? » WebNotes, [online] Available: <https://www.webnotes.com/what-are-different-types-of-search-engines/>.
- [13] How Many Types of SEO Practices Exist in the Digital World?, [online] Available: <https://www.digitalvidya.com/blog/types-of-seo/>.
- [14] 23 Search Engine Facts and Stats You Oughta Know, [online] Available: [SEO.com](https://www.seo.com).
- [15] Search Engine Use 2012, [online] Available: <https://www.pewinternet.org/2012/03/09/search-engine-use-2012/>.
- [16] Total number of Websites - Internet Live Stats, [online] Available: <https://www.internetlivestats.com/total-number-of-websites>.
- [17] Search Engine Optimization (SEO) Starter Guide - Search Console Help.
- [18] D. Bhandari, *Improving online visibility of the web pages with Search Engine Optimization*: Laurea University of Applied Sciences, 2017.
- [19] H. Banaei and A. R. Honarvar, "Web page rank estimation in search engine based on SEO parameters using machine learning techniques", *IJCSNS*, vol. 17, no. 5, pp. 95, 2017.
- [20] S. Zhang and N. Cabage, "Search engine optimization: Comparison of link building and social sharing", *Journal of Computer Information Systems*, vol. 57, no. 2, pp. 148-159, 2017.
- [21] J. Zilincan, "Search engine optimization", *CBU International Conference Proceedings...*, vol. 3, pp. 506-510, 2015.