



AUTOMATING THE MANUAL METHOD OF FINDING INDEXED BACKLINK OF WEBSITES TO IMPROVE WEB PAGE RANKING THROUGH OPTIMIZATION

Joy Bhattacharjee

Research Scholar (Computer Science and Engineering) Glocal School of Science & Technology, Glocal University, Saharanpur, Uttar Pradesh

Dr. Rajeew Yadav (Professor)

Research Supervisor, Glocal School of Science & Technology, Glocal University, Saharanpur, Uttar Pradesh

Abstract– *The reputation and trustworthy of websites depend on the quality content and good links generation. The good links generation is challenging job in diverse space of Digital Ocean. Backlinks provides the facility of good reputation of websites and ranking of page. In concern of ranking and reputation of website decide some objective related to backlines process for efficient web ranking and better search optimization. The quality of website depends on the search ability of website. Various tools and methods for off-site optimization without any cost. The process of backlinks and blog writing play an important role in quality of website. The good quality of website indicates the visibility of website over the internet. The primary objective of this paper is to propose a novel method that will reduce the cost of web marketing and increase the visibility of website over public domain and reputed search engine, such as Google, Yahoo and many more.*

I. Introduction

The digital marketing platform increase day to day due to low cost and maximum reachability of customers. The reachability of customers depends on the reputation and trustworthy of websites. The reputation and trustworthy of websites need some search optimization task for the ranking and popularity of websites. The process of search engine optimization deals with in two terms organic and inorganic. The process of inorganic SEO deals with paid services, but the service life of page rank are expiring and down after the service is closed. Instead of organic search engine optimization process is more reliable and trustworthy process of ranking website and digital content of data. To increase its rank (measured via Page Rank), intuition suggests that t could collect new in-links from j's neighbourhoods, as in real social networks, where t's new friend j introduces him to his/her acquaintances. Despite this consideration, results shown that a better rank improvement comes by acquiring long distance in-links and this occur for different networks both in type and size, therefore in this paper we exploit such evidence to build a heuristic that guarantees a target node t to achieve the best rank. Search Engine Optimization, commonly abbreviated as SEO is a series of systematic processes aimed at increasing the volume and quality of traffic through search engines to certain websites by utilizing the working mechanism or search engine algorithms. The goal of SEO is to place a website at the top position, or at least the first page of search results based on certain targeted keywords. Logically, websites that occupy the top position in search results have a greater chance of getting visitors [2]. Backlink is a link that usually shaped images or text from a particular website. If the link is clicked, then the page will directly lead to the website so that the link becomes a backlink to the website. Social Media Backlinks is a backlink that comes from social media and leads to a website in the form of text and images [3].

A backlink is created when a webpage refers to another webpage. The external link which is referring is known as Backlink. Backlink could not be neglected if someone is performing search engine optimization. Backlink holds worth when it comes to website authority. Google ranks a website based on the website authority (Google Page Rank). Google Page Rank is assigned by Google based on different website factors (Design, Visitors, and Quality of content).

In the process of organic search engine optimization “backlinks” play important role for indirect traffic generation of websites and digital content. The major concern regarding back links creation is good foundation knowledge of blog writing, effective websites content and social media link in different platforms.



Building of back links follow certain conventional guidelines for making the script of this. The process of back links formation and off-site optimization have no certain algorithm for quick response of websites. Some authors suggested that check the value of DA and PA for the performance of back links. The formation of back links and brand keys various tools and website are available. The available website cannot recognize the quality of back links and response of users after event of clicks. The major issue regarding the generation of quality data to reference the web links for the response of back links.

Therefore, automating the manual method of finding indexed backlinks of websites can significantly reduce the cost and time associated with performance improvement. By leveraging automation tools and technologies, businesses can streamline the process of identifying and analyzing backlinks, allowing for more efficient and effective SEO strategies. This transformation is critical in a competitive digital landscape where quick, data-driven decisions are essential. Manually identifying and analyzing backlinks is a time-consuming and labor-intensive task. SEO specialists need to search for backlinks, verify their quality and relevance, and track their impact on website performance. This manual process is prone to human error and can be inefficient, especially for websites with extensive backlink profiles. Automating this process can mitigate these challenges.

In current trend of research various algorithms based on network models based on connection of links for the processing of backlinks. In this synopsis we are going to propose novel machine learning based back link generation technique for the improvement of similarity of content data used heuristic based optimization techniques such as ant colony optimization and particle swarm optimization.

II. Literature Review

Several works have been carried out in this domain and still lots need to be complete on it. Various researchers have suggested their work in this domain from that some of the most significant works are discussed in this section below.

In [1]; authors present due to their high entrance barrier, the most significant work also has the highest level of competition. Consequently, the model forecasts that this work would likewise be the most hurried and of the lowest caliber. They use data from the Protein Data Bank (PDB), a database containing the structures of massive macromolecules, to evaluate the predictions of our model in the field of structural biology. The PDB's ability to assign each structure objective metrics of scientific quality is one of its key features. They find that structures with larger ex-ante potential generate more competition, are constructed faster, and are of lesser quality, as anticipated by the model. According to the model and a causal analysis of our empirical findings, these correlations are lessened when we concentrate on the structures left behind by scientists who are less focused on publication and priority due to the nature of their employment situation. This study demonstrates how the competition among researchers in the field of structural biology to publish first and claim priority leads to early publication of research, resulting in science of inferior quality.

In [2]; authors aimed at preserving competition and directing the course of inventive activities, as well as the strategic choices and lobbying strength of the companies experimenting with various technical advancements, have a significant impact during this early period. As a result, it's critical to examine how innovative businesses act in relation to working with other businesses and conducting research on certain energy solutions, as these actions represent their strategic priorities. This study uses a complex network technique to examine a network of smart grid research and development projects in order to analyse how various innovative activities branch out from these solutions and how smart grid research activities converge towards a stream of innovative solutions. The findings confirm that smart grid operations are convergent towards a single stream of innovations and identify four variants centered on distinct technological applications. The findings have implications for policymakers hoping to shape the course of smart grid research programs as well as for researchers studying technological innovation systems and smart grid innovation. They can also direct strategy development at innovative firms to engage in research collaborations or concentrate on particular smart energy technologies. The network of research projects can be properly analyzed using complex network theory to identify the primary innovation streams and the formation of variation in those streams. Such analysis's findings can assist innovative businesses in making strategic decisions about whether to collaborate on research projects or concentrate on particular smart energy technologies.



In [3]; The 4,876 public Facebook pages or groups that have shared content from the websites of 289 extremist groups listed by the Southern Poverty Law Centre (SPLC) comprise our participants, or online extremist accounts. They map the information-sharing characteristics of these extremist accounts using theories of participatory activism as a guide. Five jobs related to extreme activism are identified by clustering the quantitative variables and then obtaining qualitative expert validation: educators, solicitors, flammers, motivators, and sympathizers. For instance, while flammers spread provocative extremist content that incites rage, solicitors use links from extremist websites to beg for funds and involvement in extremist causes. They also look into role dynamics, including how long-lasting these positions are and how likely it is that extreme accounts would switch between roles.

In this work, they analyses the online ecosystem of extremist movements through the lens of participatory activism, specifically identifying five social roles in online extremist movements: educators, solicitors, flammers, motivators, and sympathizers. They found that educators and solicitors exert the most influence in triggering extremist link posts, whereas flammers are influential in triggering the spread of information from fake news sources. Additionally, they look into the dynamics and significance of roles in disseminating links from conspiratorial, biased, extreme, and fake news sites.

In [4]; authors presented a novel SIoT trust management technique that lowers the risk of exposure to malicious nodes by predicting which service providers are the most trustworthy for individual service requestors. They used a flexible bipartite graph to describe the Internet of Things, and then we use the Hellinger distance to construct a social network among the service requestor nodes. Next, we use measures of nodes' centrality and similarity to create a social trust model that extracts trust behaviours from the social network nodes. Ultimately, a matrix factorization method is developed to identify reliable nodes, extract latent information from SIoT nodes, and address issues with cold start and data sparsity. They examine how the suggested trust prediction mechanism's parameters affect the precision of the predictions. The findings show that our approach exceeds the best current techniques when it comes to receiving feedback from the neighbouring nodes of a particular service requestor with high Hellinger similarity. They also demonstrate that the prediction mechanism's accuracy is greatly increased by using the social trust model, which solely takes into account a similarity measure. They presented a revolutionary SIoT trust management mechanism in this paper. They used centrality and similarity measurements to create a social trust model that makes use of the feedback. To the best of our knowledge, this is the first study to predict trustees' trustworthiness values in the SIoT using matrix factorization. Their simulation findings showed that the malicious nodes are resisted, and our suggested mechanism accurately converges to the trustee's ground trustworthiness value. They also demonstrated the practicality of our suggested trust management approach using an actual SIoT use case.

In [5]; authors used community detection techniques, province-sector clusters that drive growth are found. Weighted PageRank scores are used to rank important provinces and sectors. A few intriguing and illuminating observations emerged from the findings. Although not as quickly as intra-province economic activity, the level of inter-province sector activities rose in tandem with the nation's rapid economic expansion. Geographical characteristics were closely linked to regional community systems. Throughout the study period, there was a rise in regional fragmentation and a significant level of community heterogeneity across the regions. The relative importance of the provinces and sectors in the national economy is measured using criteria that, in part, reflect national and regional economic development strategies.

In [6]; authors presented how normally independent ecological interactions can entwine so that the effects of an insect pest on the composition and structure of forests modify the interactions between large mammals and their predators. In particular, they demonstrate how spruce budworm irruptions, a common and cyclical insect pest of the boreal forest, modulated an indirect trophic interaction by causing moose (*Alces alces*) to benefit from a flush in deciduous vegetation. This, in turn, strengthened the apparent competition between moose and threatened boreal caribou (*Rangifer tarandus caribou*) through wolf (*Canis lupus*) predation. Crucially, salvage logging by humans intensified caribou predation after the outbreak. Our observations of large-scale, highly consequential consumer-producer-consumer interactions suggest that they are probably common in the natural world. This research shows that, particularly when human activities worsen the system, the cascading impacts of SBW in boreal ecosystems might eventually raise the mortality rate of an endangered ungulate.



In [7]; authors proposed to characterize how algorithmic curation affects topic diversity and source diversity in Twitter timelines using a sock-puppet audit. Eight sock puppet accounts were made to mimic typical real-world users who were chosen after a thorough network study. Then, in early 2020, we gathered the timelines of the puppets twice a day for a month. Overall, our findings demonstrate that while algorithmic curation significantly reduces the quantity of external links in the timeline, it enhances source variety with regard to external domains and Twitter accounts. Algorithmic curation had a mixed effect on topic diversity, suppressing clusters of tweets about health and COVID-19 fatalities while somewhat enhancing a cluster of tweets with a political bent. Lastly, we offer some proof that the timeline algorithm might make partisan disparities in their exposure to various sources and subjects worse. The results offers insight on how social media curation algorithms could deviate from conventional journalistic curation values and add significantly to the current conversation around algorithmic gatekeeping and its effects on the general audience.

In [8]; authors performed an experiment on the Facebook fan page of a Polish online retailer in a real-world commercial setting. Multiple linear regressions and metrics customised for this experiment from the literature were used to analyse the data. The findings indicate that a web link inserted into a Facebook post's comments- as opposed to the post's caption- is more profitable. It is also demonstrated that these metrics can provide useful information on the ideal posting time and the interval between postings, depending on the campaign's objectives. This would make it possible for practitioners and researchers to consider a product's quality, which was thought to be consistent in this study. There should be more research done comparing the outcomes of, say, different market niches, geographical areas, and social media platforms. Facebook's implementation of screening to limit the visibility of posts intended solely as clickbait has resulted in a minor post-publishing delay for their audiences. This hypothesis's foundation is derived from the fact that each of the 96 post's reach began extremely low and increased gradually over the course of the first few hours.

In [9]; authors analysed whether race and gender bias exist in the portrayal of artificial intelligence (AI) in picture search results from six major search engines using a mixed-method approach. According to our research, search engines favour anthropomorphic representations of AI that depict it as white, while non-western search engines are the only ones that contain non-white representations of AI. In contrast, due to increased awareness of gender bias in search results, gender representation in AI is more diversified and less biased towards any one gender. Their findings highlight the significance of developing novel strategies for identifying bias in information retrieval systems and point to the necessity and potential for resolving prejudice in the depiction of socially relevant topics, such as technological progress.

In [10]; authors present to these circumstances as "platform overthrows" and identify the pertinent tactics that a platform leader can use to thwart a challenger's plan as well as a challenger's. In order to achieve this, we reinterpreted 27 previously published instances of platform overthrow attempts using an inductive protocol. Our findings imply that both players express functional expansion and technological genericity during an attempt to topple the platform. As a result, we develop hypotheses that explain the tactics used during a platform overthrow attempt. On the other hand, functional expansion and technical genericity might potentially reveal entirely distinct information.

In [11]; authors have used Wikipedia as a computational biology instructional resource and are one of the most significant platforms for the public dissemination of research. In recent years, Wikipedia's portrayal of computational biology has significantly improved because of collaborative efforts between the International Society for Computational Biology (ISCB) and the Computational Biology taskforce of WikiProject Molecular Biology. Still, there is a pressing need for more quality improvement, particularly when compared to allied scientific domains like genetics and medicine. Encouraging participation from ISCB Communities of Special Interest (COSIs) would enhance an essential open education resource in computational biology and enable COSIs to offer a high-calibre educational resource that is highly specialised to their area of study. This research points out areas where Wikipedia's portrayal of computational biology needs work. It also shows where COSI domains are comparable, pointing to potential synergies that could strengthen linkages within the computational biology community. Lastly, we offer recommendations for future research to overcome the shortcomings of the current work and enhance an essential repository of publicly available computational biology knowledge.

In [12]; authors presented a research on competitive dynamics has primarily examined indirect factors such as the nature of the acts themselves, the enterprises engaged, or the competitive environment in order to determine



if competitors are capable of and likely to engage in competitive actions. They investigated the reasons behind competitors' initial competitive conduct at a given moment and circumstance. They used the idea of competing rationales to investigate the fundamental causes of tactical actions, drawing on the philosophy of action literature. They carried out an inductive, multicase study to investigate the rationale behind more than 800 discrete tactical decisions made by nine professional basketball coaches over the course of fifteen basketball games, given the fast-paced exchanges that define tactical competitive dynamics.

In [13]; authors shown a guarantee the web-based operations of these repositories is to assess the amount and calibre of content accessible on the websites of the Health and Medical Sciences digital repositories in North America. From Open ROAR and DOAR, a total of 72 open-access digital repositories were found. Google was the search engine used to gather data. According to the results, with 10.40% SWIF, the College of Physicians of Philadelphia Digital Library is in top place. With a smarter index value of 19.88371, Mouseion at the JAX Library holds the top spot, followed by the bioRxiv. Researchers and information professionals will benefit from this analysis, as it will give them a comprehensive understanding of the websites of this discipline's qualitative and informational repositories across North America. It implies that, in order to improve their online presence and worldwide awareness, they should implement specific methods. The current study offers comprehensive details regarding various link topologies and the online visibility of health and medical science resources. Since it is a new field of webometric research, the current study focuses on the web impact factor and link analysis of repositories across the continent of North America. Additionally, this research focuses on assessing the web presence and the linked structure of repository websites. This study offers a comprehensive overview and data regarding the websites of all 72 health and medical science repositories in the United States.

In [14]; author's major goal is to make recommendations on how these universities might raise their ranking in The League of Directors of Academic Planning's 2021 Nigerian Universities Rankings. Using the Google Scholar database, the study used a comparative methodology to evaluate the online presence of academics at three Nigerian universities. Very little online presence was found on the Google Scholar profiles of the three universities. The fact that so few academics from the three universities have Google Scholar profiles detracts from the universities' rankings. The study's findings are important not only for the three colleges in Nigeria but also for organisations in developing nations since, if put into practice, they will raise rankings and, in turn, prospective students will find the university more appealing. The study's conclusions will advance knowledge of the necessity for academics to build their online presence, which may have an impact on their university rankings. They will also make academics more conscious of the need to verify the citations for their work.

In [15]; author's study is an innovative attempt to investigate the websites of worldwide library associations and evaluate those websites using webometric metrics and web content evaluation. Researchers have examined a number of library websites. On the other hand, no such endeavour has been made in relation to the website of the library organisation on a global scale. The sample consisted of international library associations that were indexed on the website of the American Library Association. Google was used for the linkages, and data were gathered using a link analysis tool from <https://moz.com/> that aids in assessing domain authority, page authority, in links, and out links. The official library association websites were examined using a well-organised checklist consisting of 19 criteria. It has been discovered that the International Association of Law Libraries has the most external and easy web impact factors. The Association of Jewish Libraries, on the other hand, has the greatest internal web impact factor. Every website that was looked at for this study is housed under the.org domain.

In [16]; authors presents a news media firms generate their own material and are paid for it through advertising and subscriptions, search engines that choose and display content that they did not create themselves sometimes may not have to pay copyright fees. They examined how the two linking approaches impact the technical efficiency of media firms, focusing on search engines that employ them to provide news material to users. However, there isn't much empirical research on how news media firms' performance efficiency varies based on how search engines deliver news material. Using a meta-frontier production function analysis, the findings indicate that media firms, irrespective of their size, financial standing, or reputation, face significant constraints from search engines when offering news services via internal connections. This study concentrated on the intense disagreement that exists between search engines and news media companies as a result of the rivalry between new and established technological enterprises.



In [17]; authors looked at affiliation differences between Microsoft Academic, Web of Science (WoS), Dimensions, and Scopus for 18 Arab universities. They verified that digital object identifiers, or DOIs, are appropriate for extracting similar academic content from different databases and measured the differences in affiliation between them. In each given database, a significant portion of the entries assigned to the chosen universities were not assigned to the same university in another. The larger datasets have a higher discrepancy share. A higher degree of agreement was seen between affiliations in the smaller, more selective databases and those in the other databases. A manual analysis of the affiliation disparities revealed that a combination of missing affiliations, variations in unification, and incorrect institution assignment was the root reason. In our sample, they discovered that variations in unified variants were typically the cause of differences in WoS, and in certain instances, it was evident that uncertainty resulted in data being assigned to the incorrect institution.

In [18]; authors examined how a family business used its three generations of entrepreneurial experience to sustain profitable expansion over a 16-year period, using a rhetorical history approach. Their analysis reveals three history scripting strategies: constructing, elaborating, and embedding family history. These strategies gave the family business a significant competitive advantage, aided in its acceptance by larger communities, established its reputation for continuity, and encouraged innovation through tradition. For family businesses, history is a significant source of competitive advantage. By using a rhetorical history viewpoint, we have established and introduced the idea of scripted history in this study, exposing several history scripting techniques that explain the strategic uses of history to provide long-term competitive advantage. They do not assert that the list of competitive advantages resulting from an entrepreneurial family's past is all-inclusive in this study.

In [19]; authors presented a popular method for producing small-molecule protein inhibitors and treatment candidates is fragment-based drug discovery, which is employed by both academic institutions and pharmaceutical businesses. The connecting strategy is one of the methods described in the literature that supposedly provides a quick increase in binding energy. However, when looking at the drugs that are presently undergoing clinical trials, this strategy is not well represented. Here, they present a thorough analysis of all the cases that have been documented in the literature to date, along with the techniques employed to concurrently or sequentially identify the two starting fragments. Combining linking and merging techniques, in which the linker is copied from a recognised inhibitor, seems like a promising tactic. Similar to the compounds where super-additivity was attained, the linkers are primarily formed from alkyls and are rather short and flexible.

In [20]; authors shows people frequently use the Internet to research scientific topics when they have queries. In the post-truth era, evaluating the sources of information and assertions has become increasingly challenging. Although source evaluation skills are frequently taught to students, the spread of "fake news" has led to an arms race in misinformation. The ability of searchers to recognise deceptive material is growing, and with it, so too do the information providers who wish to mislead. Based on a theoretical framework of plausibility judgements as well as contemporary theories and research on source evaluation, we propose that the post-truth era has increased the importance of critically analysing online material related to scientific matters. In spite of these obstacles, the post-truth era has led to a heightened awareness of the necessity of critically assessing online content about scientific matters. Developments in epistemic cognition are making evaluation techniques that could make online users of scientific knowledge more acutely aware. We contended that analysing the relationships between information sources and knowledge assertions may require explicitly evaluating plausibility judgements.

In [21]; authors shown a knowledge graph personalised search model, or KEPS, is presented in this study. KEPS creates a knowledge-enhanced profile for each user and her searches by first performing personalised entity linking on the queries and creating better intent representations; it then uses memory networks to store the predicted search intents and connected entities in her search history. KEPS then uses the intent representation and knowledge-enhanced user profile to provide a more personalised, knowledge-enhanced search. Furthermore, KEPS uses user feedback to post-adjust the entity linkage on earlier inquiries after the provision of personalised search for each query. By doing this, past linking mistakes are fixed, and future query rankings are of higher quality. We intend to modify connection probabilities in the future by modelling the relationship between things using graph neural networks.

In [22]; presents an examining potential relationship between user-driven factors, user ratings, and ranking algorithms was the aim of this correlational quantitative study. Students and teachers in the information



technology (IT) department at a university in Huntington, West Virginia, made up the study's population. The impossibility theorem of Arrow served as the theoretical foundation for this investigation. In order to determine the correlations between the variables, a multiple regression analysis was performed on the whole survey data that was gathered from 47 IT faculty members and students. Eighty-five percent of the ranking algorithm's variability could be explained by the model.

In [23]; authors provide a method for creating event-centric sub-collections from such massive archives, which comprise, more crucially, documents explaining associated elements in addition to the core documents relevant to the event itself. This is accomplished by first extracting pertinent concepts and entities from a knowledge base and then looking for mentions of those concepts and entities in texts, which are then interpreted as relevance indicators. We thoroughly assess our system on two diachronic corpora: the US Congressional Record and the New York Times Corpus. In addition, we examine its performance on two publicly available large-scale web collections: the TREC-CAR dataset and the TREC KBA Stream Corpus. The most effective single approach enlarges the query by embedding representations of pertinent entities and event characteristics. This method can sometimes already get good results, depending on the type of event and the collection.

In [24]; this study, they can gather vulnerability information based on huge amounts of data that is already present on the network by building a cyber-attack framework based on preemptive prediction, and we can improve the accuracy by applying machine learning to the keyword mapping. The US Department of Defence's and BAS's CHESS project, which is currently in the commercialization stage, is one of the better examples. An automated simulation of attack scenarios is carried out by BAS to identify vulnerabilities in the current environment and simulate several attack scenarios for a more accurate vulnerability analysis.

In [25]; authors refers Intrusion detection systems (IDSs) have become essential for protecting organisational and personal assets due to the rapid increase in cyberattacks. But as computer networks get bigger, it gets harder for a stand-alone intrusion detection system (IDS) to recognise complex and advanced threats like denial-of-service attacks because it doesn't have enough contextual data or understanding of the contexts in which it is being deployed. Distributed and collaborative intrusion detection systems (IDSs) are designed to address this problem. They allow a group of IDS nodes to work together by sharing necessary data. In this review, they first address the primary external and internal threats and provide an overview of the state-of-the-art for traditional DIDSs based on the collaborative topology (centralised, decentralised, and distributed). Trust is frequently established among different IDS nodes due to the scattered nature and variety of threats. Next, they present a chronological summary of the pertinent studies on trust-based DIDS. They determine that the emergence and growth of the Internet of Things may present more opportunities as well as obstacles for trust-based CIDNs.

In [26]; authors discusses Search Engine Optimization (SEO) plays a very vital role in the development of professional websites. There are some search engines available on the internet such as Yahoo, Ask.com, AOL.com, Baidu, and Bing. Among which Google is the most widely used search engine. Each search engine uses different SEO technique and algorithm, which not only forms the foundation of SEO but affects the position of a website in organic search results as well. As Google modify its algorithm about 500 or more times per year, the web design and internet also evolve dynamically because of changes in SEO techniques and algorithms. However, how well Malaysian universities websites are optimized for other search engines is questionable particularly the critical differences between search engine ranking techniques and algorithms. This research paper tends to answer these vital questions by proposing a comparative analysis of Bing and Google on some Malaysian university's website, analyzing their search engine optimization parameters and outcomes of using Microsoft Bing as compared to its primary competitor, Google.

In [27]; authors evaluate the content and technical characteristics that influence webometric rankings on Golestan University of Medical Sciences websites. This research article is a descriptive practical survey that describes the status of the website based on the content and technical features that influence the promotion of webometric rankings. A total of 53 Golestan University of Medical Sciences websites were included, consisted of 7 college websites, 8 Deputies websites, and 16 websites for hospitals and 22 websites for research centers. Data were collected using checklist. Result shows that 49.46 percent of the expected indicators were met by the units. Colleges with the overall average of 57.64 percent had the highest compliance and the hospitals with 36.72 percent had the lowest compliance with the webometric indexes. The results show that 54.72 percent of the Deputies and 47.73 percent of the research centers have met the indicators.

In [28]; they discussed a method of phishing website detection that utilizes a meta-heuristic-based nonlinear regression algorithm together with a feature selection approach. In order to validate the discussed method, they used a dataset comprised of 11055 phishing and legitimate webpages, and select 20 features to be extracted from the mentioned websites. This research utilizes two feature selection methods: decision tree and wrapper to select the best feature subset, while the latter incurred the detection accuracy rate as high as 96.32%. After the feature selection process, two meta-heuristic algorithms are successfully implemented to predict and detect the fraudulent websites: harmony search (HS) which was deployed based on nonlinear regression technique and support vector machine (SVM). The nonlinear regression approach was used to classify the websites, where the parameters of the discussed regression model were obtained using HS algorithm. The discussed HS algorithm uses dynamic pitch adjustment rate and generated new harmony. The nonlinear regression based on HS led to accuracy rates of 94.13 and 92.80% for train and test processes, respectively. As a result, the study finds that the nonlinear regression-based HS results in better performance compared to SVM.

In [29]; authors presents the most of the times the main focus is set on making the website as user-friendly as possible, stable, fast, and secure. However, all of these techniques can be useless if the website does not have any visitor or people just cannot find it. To solve this problem, and to enhance the structure of the website to become more search-engine-friendly, a web application is developed to analyze any given webpage and to provide information on how to enhance and improve its structure to make it more search-engine-friendly and to improve its ranking on search engines. This process is also known as Search Engine Optimization (SEO). On-page SEO and off-page SEO are both techniques used to improve the visibility of a website. In addition to those two techniques, social media is a very important aspect of search engine optimization (SEO), because it helps driving more targeted visitors to the website.

In [30]; this article, authors carried out Multiple Linear Regression Analysis to identify the effect of domain age and Google indexed pages on total monthly page views. They used the data gathered through web analysis tools such as SimilarWeb, SmallSeoTools and Pingler on 60 websites from India accessible to us. They deployed ANOVA to explore the differences in average paid search traffic, social media traffic, mail traffic, referrals traffic and display traffic with respect to different website categories. The results thus obtained clearly demonstrate that domain age and Google indexed pages have a weak impact on website traffic but there is a significant difference in average organic search traffic, paid search traffic, social media traffic and referrals traffic with respect to those website categories. This article analyzed the relationship of monthly page views with domain age and Google indexed pages while assessing the traffic flow behavior based on website categories. The results show that Google indexed pages have a statistically significant relationship on monthly page views with a low to moderate strength. However, the relationship between domain age and monthly page views was found insignificant. This relationship can also be influenced by extraneous variables such as category of website, quality of content and others. Though the relationship between Google indexed pages and website traffic is significant, it shows only weak to moderate strength. This proves that there could be other factors that may influence website traffic further, deeper or stronger. Or, the website traffic can be influenced by Google indexed pages when it is combined with other major factors such as backlinks, Bing indexed pages and content quality.

In [31]; authors have been used an auto search tool to increase the ranking of their websites automatically. It is efficient to get the Result in a timely manner. Two approaches were inferred; firstly, selection of keywords and deep analysis method in SEO is applied to deal with the problems like low speed, deprived reliability and not flexible optimization in SEO processing. Secondly, Web Meta Data (WMD) is applying to real situation using different approaches to deal with low visibility in Search Engine. Traffic generated by this tool is more as compared to real traffic. In this study, the auto search tool can increase the web site ranking automatically. Traffic generated by this tool is more as compared to real traffic. Search Engine Optimization could be enhanced effectively and it can produce better benefits for all those stakeholders. The increase in number of visitors results in increase in ranking. In this study, the auto search tool can increase the web site ranking automatically. Traffic generated by this tool is more as compared to real traffic. Like, there are 50 visitors per day and set auto search ratio = 150% and when tool is applied $50 * 1.5 = 75$ visitors are added. Tool traffic will be mixed with real traffic and could look natural. In this study, the auto search tool can increase the web site ranking automatically. Traffic generated by this tool is more as compared to real traffic.

III. Proposed Methodology

One of the primary outcomes of the proposed model is an improvement in search engine rankings. By acquiring high-quality backlinks, the algorithm boosts the website's domain authority, making it more likely to rank higher for relevant search queries. This increase in visibility drives more organic traffic to the website, enhancing its overall branding and reach. The proposed algorithm also positively impacts user engagement. Proposed methodology of backlinks optimization is shown in figure-1 below.

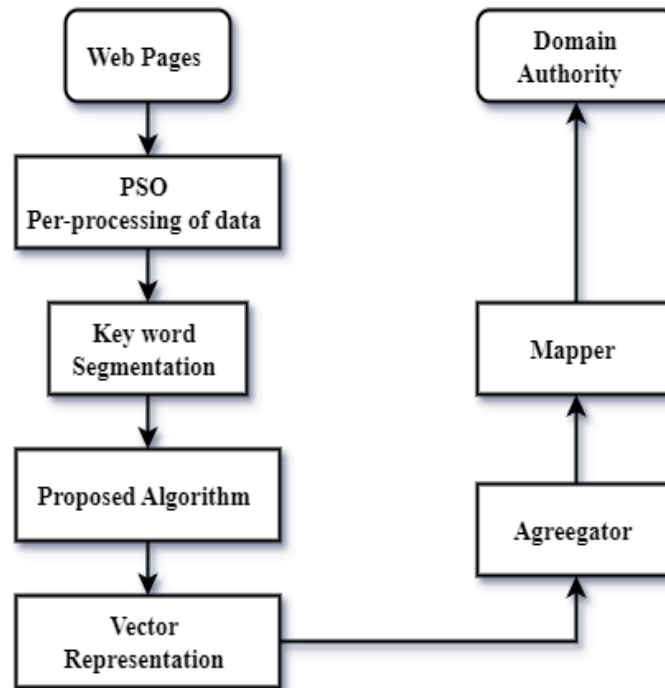


Fig. 1: Proposed Model of Backlinks Optimization

3.1 Problem Domain

The process of backlinks generation and mapping is major task of off-site optimization. The quality of backlinks creation is good the performance of website is automatically improved. The major issue of backlines creation is developed network connection of data redirection. Some authors used the methods of time-based filter design for branding of keywords for the improvement of searching capacity of backlinks. Some problems and issues related to backlines mention here.

- The brand key links in web contents
- The similar links creates spam backlines
- The efficient content management and distribution of content
- Ensemble based design network not cover the all-web networks

3.2 Objectives of the Research

Following are the main Objectives of the proposed research as under-

- Design the content filter-based algorithms for branding of keywords
- Design machine learning based algorithm for network generation of links
- Modelling of data for creation of backlinks
- Improve the performance of DA (domain authority) and PA (page authority)

- Reduces the cost of search engine optimization.

3.3 Data Sets

We will use Freebase dumped in January 2014 as our knowledge base. Freebase classifies entities into types under various domains such as people/person and book/author. An entity might belong to multiple types in Freebase. Each type under a domain is treated as a query type in our research work. We then will remove very sparse types such that each type has at least five entities. Finally, there are 444 query types in our type set.

IV. Conclusion

This research study provides an in-depth analysis of current methods and tools for finding and analysing indexed backlinks. Performing manual backlink analysis on massive amounts of web data is a tedious, error-prone, and wasteful process. Automating the process of finding and evaluating backlinks is becoming increasingly important for Search Engine Optimisation (SEO) tactics, according to the report. The article showed how automation may make backlink research much faster, more accurate, and scalable by examining optimisation algorithms, data extraction methods, and automated backlink indexing strategies. Web page ranking and online visibility are both enhanced by the incorporation of optimisation strategies, which allow for a more accurate evaluation of backlink quality, relevance, and indexing status. In conclusion, the research proves that replacing human link discovery with an automated approach is an important first step in developing data-driven SEO systems that can adjust to ever-changing search engine algorithms. The future scope of this research is broad and present numerous avenues for additional investigation. Sophisticated machine learning and artificial intelligence methodologies can be integrated to more precisely forecast backlink quality, spam probability, and the long-term influence on search rankings. Real-time backlink monitoring systems can be established to automatically detect variations in indexing status and notify website administrators.

Furthermore, these developments will enhance robust, intelligent, and ethical SEO tactics, benefiting both website proprietors and search engine ecosystems.

References:

- [1]. Hill, Ryan, and Carolyn Stein. "Race to the bottom: Competition and quality in science." *Northwestern University and UC Berkeley* (2021).
- [2]. Dehdarian, Amin, and Christopher L. Tucci. "A complex network approach for analyzing early evolution of smart grid innovations in Europe." *Applied Energy* 298 (2021): 117143.
- [3]. Phadke, Shruti, and Tanushree Mitra. "Educators, solicitors, flammers, motivators, sympathizers: characterizing roles in online extremist movements." *Proceedings of the ACM on Human-computer Interaction* 5, no. CSCW2 (2021): 1-35.
- [4]. Aalibagi, Soroush, Hamidreza Mahyar, Ali Movaghar, and H. Eugene Stanley. "A matrix factorization model for hellinger-based trust management in social internet of things." *IEEE transactions on dependable and secure computing* 19, no. 4 (2021): 2274-2285.
- [5]. Wang, Tao, Shiyang Xiao, Jun Yan, and Panpan Zhang. "Regional and sectoral structures of the Chinese economy: A network perspective from multi-regional input-output tables." *Physica A: Statistical Mechanics and Its Applications* 581 (2021): 126196.
- [6]. Labadie, Guillemette, Philip D. McLoughlin, Mark Hebblewhite, and Daniel Fortin. "Insect-mediated apparent competition between mammals in a boreal food web." *Proceedings of the National Academy of Sciences* 118, no. 30 (2021): e2022892118.
- [7]. Bandy, Jack, and Nicholas Diakopoulos. "More accounts, fewer links: How algorithmic curation impacts media exposure in Twitter timelines." *Proceedings of the ACM on human-computer interaction* 5, no. CSCW1 (2021): 1-28.
- [8]. Chawla, Yash, and Grzegorz Chodak. "Social media marketing for businesses: Organic promotions of web-links on Facebook." *Journal of Business Research* 135 (2021): 49-65.

- [9]. Makhortykh, Mykola, Aleksandra Urman, and Roberto Ulloa. "Detecting race and gender bias in visual representation of AI on web search engines." In *International Workshop on Algorithmic Bias in Search and Recommendation*, pp. 36-50. Cham: Springer International Publishing, 2021.
- [10]. Thomas, Maxime, Pascal Le Masson, Benoît Weil, and Julien Legrand. "The future of digital platforms: Conditions of platform overthrow." *Creativity and Innovation Management* 30, no. 1 (2021): 80-95.
- [11]. Kilpatrick, Alastair M., Farzana Rahman, Audra Anjum, Sayane Shome, KM Salim Andalib, Shrabonti Banik, Sanjana F. Chowdhury et al. "Characterizing domain-specific open educational resources by linking ISCB Communities of Special Interest to Wikipedia." *Bioinformatics* 38, no. Supplement_1 (2022): i19-i27.
- [12]. Andreovski, Goce, Danny Miller, Isabelle Le Breton-Miller, and Walter Ferrier. "Competitive rationales: Beneath the surface of competitive behavior." *Journal of Management* 48, no. 8 (2022): 2286-2317.
- [13]. Ghosh, Sanjib, and Bijan Kumar Roy. "Webometric analysis of open access repositories of health and medical sciences in the continent of North America." *Library herald* 60, no. 1 (2022): 148-170.
- [14]. Idiedo, Victor, and Clement Okeji. "An Investigation into the Web Presence of Academic Staff of Universities in Nigeria: Impact on University Ranking." *Niger Delta Journal of Library and Information Science* 3, no. 1 (2022).
- [15]. Dhar, Prasenjit, and Mithu Anjali Gayan. "A webometric study of selected international library association websites: An evaluative study." *DESIDOC Journal of Library & Information Technology* 42, no. 3 (2022): 185-190.
- [16]. Cho, Yongsun, Daeho Lee, and Inyoung Park. "Comparison of Media Company Performance Efficiency Based on the Search Engine's Method of Providing News Content (External Links vs. Internal Links)." *Applied Sciences* 12, no. 23 (2022): 12327.
- [17]. Purnell, Philip J. "The prevalence and impact of university affiliation discrepancies between four bibliographic databases—Scopus, Web of Science, Dimensions, and Microsoft Academic." *Quantitative Science Studies* 3, no. 1 (2022): 99-121.
- [18]. Ge, Bingbing, Alfredo De Massis, and Josip Kotlar. "Mining the past: History scripting strategies and competitive advantage in a family business." *Entrepreneurship Theory and Practice* 46, no. 1 (2022): 223-251.
- [19]. Bancet, Alexandre, Claire Raingeval, Thierry Lomberget, Marc Le Borgne, Jean-François Guichou, and Isabelle Krimm. "Fragment linking strategies for structure-based drug design." *Journal of Medicinal Chemistry* 63, no. 20 (2020): 11420-11435.
- [20]. Sinatra, Gale M., and Doug Lombardi. "Evaluating sources of scientific evidence and claims in the post-truth era may require reappraising plausibility judgments." *Educational Psychologist* 55, no. 3 (2020): 120-131.
- [21]. Lu, Shuqi, Zhicheng Dou, Chenyan Xiong, Xiaojie Wang, and Ji-Rong Wen. "Knowledge enhanced personalized search." In *Proceedings of the 43rd International ACM SIGIR conference on research and development in information retrieval*, pp. 709-718. 2020.
- [22]. Taylor, Gary Michael. "Search Results: Predicting Ranking Algorithms With User Ratings and User-Driven Data." PhD diss., Walden University, 2020.
- [23]. Nanni, Federico, Simone Paolo Ponzetto, and Laura Dietz. "Toward comprehensive event collections." *International Journal on Digital Libraries* 21 (2020): 215-229.
- [24]. Ryu, S., J. Kim, N. Park, and Y. Seo. "Preemptive Prediction-Based Automated Cyberattack Framework Modeling. Symmetry 2021, 13, 793." (2021).
- [25]. Li, Wenjuan, Weizhi Meng, and Lam For Kwok. "Surveying trust-based collaborative intrusion detection: state-of-the-art, challenges and future directions." *IEEE Communications Surveys & Tutorials* 24, no. 1 (2021): 280-305.
- [26]. Shahzad, Asim, Nazri Mohd Nawi, Edi Sutoyo, Muhammad Naeem, Arif Ullah, Sundas Naqeeb, and Muhammad Aamir. "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, no. 4 (2018): 1262-1269.
- [27]. Mohammadi, Masoud, Roqiyeh Saeedi, Mohsen Mansouri, Maryam Banisaffar, Amir Mahimani, Mojtaba Azghandi Shahri, and Mohammad Yadgari. "Evaluation of Effective Indicators on Promotion of Webometric Rank of Golestan University of Medical Sciences Website." *Webology* 16, no. 2 (2019): 242-255.
- [28]. Babagoli, Mehdi, Mohammad Pourmahmood Aghababa, and Vahid Solouk. "Heuristic nonlinear regression strategy for detecting phishing websites." *Soft Computing* 23, no. 12 (2019): 4315-4327.
- [29]. Barbar, Aziz, and Anis Ismail. "Search Engine Optimization (SEO) for Websites." In *Proceedings of the 2019 5th International Conference on Computer and Technology Applications*, pp. 51-55. 2019.
- [30]. Chakraborty, Kalpita, and Ebbin Jose. "Relationship Analysis between Website Traffic, Domain Age and Google Indexed Pages of E-commerce Websites." *IIM Kozhikode Society & Management Review* 7, no. 2 (2018): 171-177.
- [31]. Zahoor, Mehak, and Shafaq Zahoor. "Comparative Analysis of Search Engine Optimization Using Auto Search Tool."

