Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Hybrid MA-BBO-PSO Metaheuristic Framework for Large-Scale Website Reorganization and Enhanced Web Navigation

Ravi Kumar

ravikumar_08@outlook.com

ABES Engineering College

Santosh Kumar

Galgotias University

Research Article

Keywords: Biogeography-Based Optimization, Particle Swarm Optimization, Memetic Algorithm, Hybrid Metaheuristic Algorithm, Webpage Ranking, Website Reorganization, and Information Retrieval

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10354860/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Hybrid MA-BBO-PSO Metaheuristic Framework for Large-Scale Website Reorganization and Enhanced Web Navigation

Ravi Kumar[1*], Santosh Kumar[2]

[1]Department of Computer Science and Engineering, ABES Engineering College, Ghaziabad, Uttar Pradesh, India-201009

[2]School of Computing Science and Engineering, Galgotias University, Greater Noida, India

[*]Corresponding Author Email: ravikumar_08@outlook.com

Abstract

A high rate of growth of web material has made navigating and retrieving information more difficult. To rank Web pages higher and dynamically reorganize large websites, this work proposes a new hybrid metaheuristic algorithm (WDR_MA_BBO_PSO) that is a combination of Memetic method (MA), Biogeography-Based Optimization (BBO) and Particle Swarm Optimization (PSO). Webpages are ranked based on a number of behavioral and structural criteria including the frequency of the keyword used, the frequency of access, the length of the session and the hub-authority rating. Experimental validation of the suggested algorithm on both synthetic and real-world data shows that it is superior in comparison to such traditional algorithms as GA, SA, PSO, and PageRank. The WDR_MA_BBO_PSO algorithm with a precision of 92, recall of 79, and average CTR value of 39 exhibited a good convergence with a change in population size. The method reduces the search depth and enhances user interaction by placing pages where a user rates them high nearer to the home page. This is an option that offers scalability of restructuring websites to adapt to a wide range of evolving user requirements.

Keywords: Biogeography-Based Optimization, Particle Swarm Optimization, Memetic Algorithm, Hybrid Metaheuristic Algorithm, Webpage Ranking, Website Reorganization, and Information Retrieval.

1. Introduction

The internet is now too huge with so many web pages and one can hardly find the necessary information. Although it is their best efforts to produce something due to user queries, the existing search engines often end up showing the least relevant stuff even when it is already stored in the system (Ramya & Shreedhara, 2017) .One of the leading causes of this inefficiency is the limitations of the traditional information retrieval (IR) methods that are unable to answer real-time queries because of the nature and volume of web data, which are too voluminous and complicated (Bhopale and Tiwari, 2020) .

The conventional methods of IR are becoming less flexible, more exact and more realistic with the web information ever-growing in quantity and variety. It has been proposed that many approaches can be used to optimize document retrieval, such as Relevance Feedback (Wang et al., 2021), Genetic Algorithms, and Evolutionary Computational approaches, including Ant Colony Optimization (Irfan and Ghosh, 2017) and Particle Swarm Optimization

(PSO). Such methods often produce materials which are partly or wholly useful, although they are trying to increase search effectiveness.

The optimization of web information retrieval (WIR) has become the key to eliminating these challenges.The fact that PSO has been successful in its capacity to emulate social behavior so that it can explore large search areas is very well known (Drias, 2011). There are also improved techniques such as PSO1-IR and PSO2-IR which have proven to be more accurate and scalable in benchmark sets such as CACM and RCV1. Metaheuristic algorithm used to enhance web navigation by recombining large scale website structures. In contrast to the past where researchers mainly focus on one form of data either content, usage, or structure, we take into consideration several variables to measure the significance of webpages and they are the time spent on pages, unique visitors, frequency of access, frequency of key words in terms of previous navigation, linkage metrics such as hubs and authority scores.

The proposed hybrid algorithm includes this multidimensional data to rank the webpages based on Melodic method (MA), Biogeography-Based Optimization (BBO), and Particle Swarm Optimization (PSO) to find out the most relevant pages. When using huge sites, it is possible to enhance the overall navigability and the user experience by ensuring that these highranking pages are brought as near to the home page as possible thereby making the search depth shallow.

This paper demonstrates that multi-criteria assessment can significantly enhance the relevance and availability of web content besides providing a new hybrid metaheuristic model of web site re-structuring.

1.1 Metaheuristic Optimization Algorithms (MOA)

It is widely recognized that the Metaheuristic Optimization Algorithms (MOA) is capable of giving somewhere near the optimal solution of the complex global optimization problems that would otherwise be incredibly computational ineffective to solve using precise algorithms.

One of the applications that requires the management of a large and dynamic search area is known as ranking and reorganization of web documents.

MOAs are known to be versatile, flexible and have the ability to search the world wide as well as are often manipulated by social or environmental factors. Some of the algorithms have been successfully applied in web mining, feature selection, clustering, and user behavior modeling. These are Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Simulation Annealing (SA), Biogeography-Based Optimization (BBO) and Memetic Algorithms (MA).

In Figure 1, a taxonomy of significant metaheuristic strategies that can be applied in this study is provided.

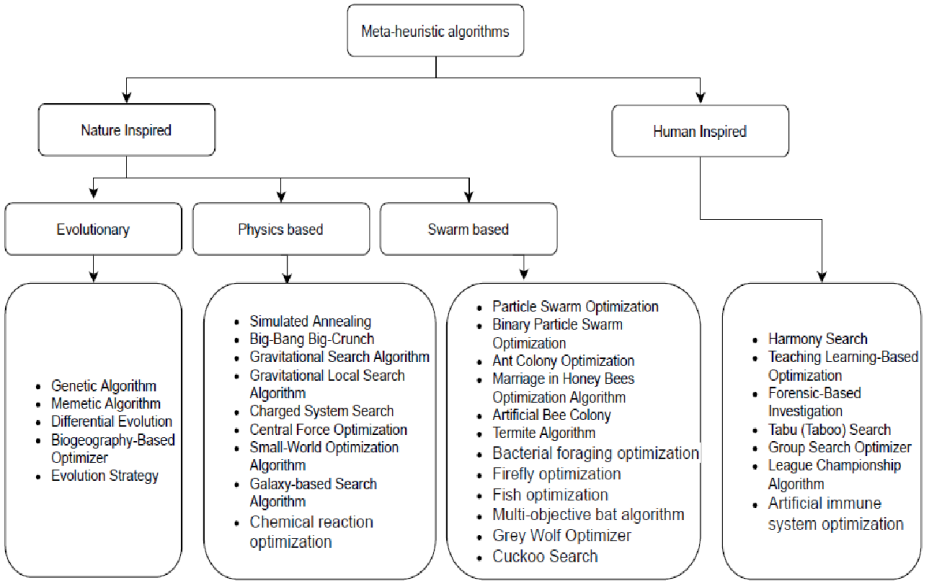


Figure. 1. Taxonomy of Metaheuristic Optimization Algorithms

2. Related Work

Since the web content is growing exponentially, information retrieval (IR) has grown to be a significant field of research with the objective of gathering relevant items based on the query by the user. Sometimes early methods of IR provided insufficient or inappropriate results as they could not capture context and intent of the user as they relied on matching of documents based on the frequency of keywords.

2.1 Structured Review

2.1.1 Classical and Vector-Based Information Retrieval

Ren et al. (Ren 2013) provided a document retrieval method based on the Vector Space Model. They are discrete stochastic optimization-based algorithm, including a maximum similarity criterion and having self-learning abilities to improve the efficiency of calculations. To enhance the relevancy in responses, El-Gayar et al. [6] subsequently enhanced conventional methods of keyword frequency by integrating the semantic indexing and interrogative context consideration in queries.

2.1.2 Genetic and Evolutionary Computation Approaches

The genetic algorithms (GAs) have been used to do a lot of work on IR difficulties. Fan et al. (Fan et al. 2009) presented a GA-based architecture in two steps. The former is based on the use of HTML content to determine the statistical trends and the latter is based on different ranking techniques to optimize the performance. This was also the case with Sharma et al. (Sharma et al. 2024) who developed a GA-based inquiry optimization algorithm that would generate higher retrieval precision than three previously existing GA-based models on benchmark datasets (CACM, Cranfield, and Medline).

Bessai-Mechmache and Hammouche [not numbered above] applied genetic algorithms in order to retrieve XML information by reducing semantic distance between user queries and XML objects. Shadab Irfan and Ghosh also emphasized the importance of evolutionary computation when they combined PSO, GA, and Ant Colony Optimization to give strong results in the process of retrieval.

2.1.3 Swarm Intelligence and Metaheuristics

Swarm intelligence methods have proved to be highly promising in IR:

- Drias et al. (Drias 2011) developed both sequential and parallel Particle Swarm Optimization (PSO) algorithms of web IR. Their test outcomes on CACM and RCV1 data sets established that parallel PSO was superior compared to classical methods in terms of quality and performance (time) of the solutions.
- Various methods have been proposed by Djenouri et al. (Djenouri et al. 2018) as a Bees Swarm Optimization technique, in which preprocessing was performed with K-means and DCI-Closed clustering before being explored with swarm methods. The quality of the document retrieval on their method was high on the datasets such as TREC, CACM and Wikileaks.
- Zhao et al. (Zhao et al. 2022) and Singh et al. (Singh and Kaur 2021) used Ant Colony Optimization (ACO) to rearrange websites. Zhao suggested to cut on search depth and computations time with the help of ACO and Singh presented the problem of webpage arrangement as a Quadratic Assignment Problem (QAP) and compared standard ACO with ACO with heuristics.

2.1.4 Website Structure Optimization

There are many heuristic methods explored in optimization of the structure of websites:

- Yin et al. (Kumar et al. 2023) suggested an Enhanced Tabu Search (ETS) algorithm, which gave the opportunity to use cyclic paths to estimate bidirectional associations. ETS was more efficient in restructuring as compared to GA.
- Considering user preferences, display time and the existing trends, Panicker et al. (Panicker et al. 2024) proposed a flexible way of reconstruction of e-commerce websites.
- To detect violations and propose structural changes, Bessghaie et al. (Bessghaier et al. 2022) focused on the structural design of UI by relying on the framework that was developed based on the material design principles of Google.

2.1.5 Semantic and Contextual Retrieval Techniques

Scholars examined contextual and semantic models to fulfill the intent of a user.

- Kim et al. (Kim 2019) provided a burst state detection model that emphasizes on ranking of pages based on user interest and ongoing events and rearranging of search results in changing circumstances e.g. natural disasters.
- Jin et al. (Jin et al. 2017) proposed WebPagePrev, a re-visiting model that makes use of memory context-based and content-based memory. It was much better in F1-score and re-finding accuracy than current algorithms such as Memento and search history lists.

2.1.6 Link-Based Ranking Algorithms

Link analysis has played a very important role in web page ranking:

- So as to locate hubs and specialists in topic searches, Kleinberg (Jalilifard et al. 2021) created HITS (Hyperlink-Induced Topic Search).

- By Brin and Page (Gampa and Fujita 2021), web search was changed by the introduction of PageRank, which returns better ranking of pages based on their link structure.
- Diligenti et al. (Diligenti et al. 2004) studied such methods as probabilistic scoring systems, HITS, and WPR as far as page ranking is concerned. They highlighted the problem of indexing and proposed hybrid schemes, which would integrate the vertical (content-based) and horizontal (link-topology) search paradigms.

2.1.7 TF-IDF and Hybrid Scoring Models

- Ramos et al. (Oleinik 2022) used the TF-IDF model with document length normalization, which led to the higher retrieval accuracy in diverse datasets.
- Using relevance feedback, Amati et al. (Bessghaier et al. 2022) proposed to the BM25 a set of improvements that would enable the model to extract knowledge due to user interaction.
- The hybrid system that entailed RankBrain and BM25 was studied by Gampa et al. (Gampa and Fujita 2021), and it demonstrated the improved understanding of the user query semantics and document content.

2.1.8 Domain-Specific and Regional Studies.

- Vyas et al. (Vyas 2019) examined the websites of the Indian state tourism and focused on the lack of past research by users on Indian government web ranking in terms of foreign visitor behavior using SEO tools such as Google Trends, Alexa and SimilarWeb.

- In evaluating IR methodologies, Mathee et al. (Abdulmunem and Naamha 2023) found out that the Content Rank and Weighted PageRank are widely used in analysing the linkages in WIR systems.

This systematic review of the existing literature indicates the shift in the traditional methods of retrieval of information by using keywords to more sophisticated methods of multi-objective and context-sensitive optimization. However, most of the approaches only focus on structure, use or contents. This inconsistency is the motivation of the Hybrid MA-BBO-PSO Metaheuristic that binds all of these three in a multi-criteria framework to enhance the navigability and quality of ranking of large-scaled web sites.

2.2 Literature Gap

Most of the modern approaches consider only a part of the page ranking features, including key word frequency, link structure or number of visits, even though there has been a great amount of research on the areas of information retrieval (IR) and website structure optimization. The few researches have not incorporated more than a single factors in a single ranking model like the usage patterns, content semantics and structural relevance. The homogenous nature of the models that are being applied today does not take into consideration the different tastes and information needs of the different categories of users. Consequently, the happiness of the consumers is diminished and the personalization is not optimal.

The majority of systems of website reorganization do not utilize behavioral data such as clickstream trends, session lengths, scroll depth, and heatmaps. These data sources will be able to give valuable information in the development of navigation paths more user centered. The algorithms such as Genetic Algorithms (GA) and PSO have been studied individually but hybrid metaheuristic Methods have not been well studied. Also, it has no integration with

machine learning or AI-based decision-making, particularly in the large-scale real-time web settings.

2.3 *Proposed Contribution*

This hybrid approach provides a scalable and adaptable method of reorganising large websites without losing performance and relevance to the changing needs of the users.
To overcome these limitations, the present paper will suggest a new Webpage Dynamic Reorganization (WDR_MA_BBO_PSO) method. It achieves this through:

- a multi-criteria ranking model which involves the use of links, content and usage metrics.
- Able to dynamically modify the structure of the web pages depending on the rating of pages to enhance navigability and user experience.

It is a hybrid mainstream that utilizes the advantages of Particle Swarm Optimization (PSO), Biogeography-Based Optimization (BBO) and Memetic Algorithm (MA). It proves to be a better search space exploration and perfect webpage ranking following careful analysis.
This hybrid solution offers a scalable and flexible procedure of reorganizing large websites without impacting performance and relevancy in response to the changing needs of customers.

3. Proposed Framework

3.1 Problem Formulation

The issue of re-arranging the web structure is also of paramount concern as regards to the business development perspective. Any reputable site, usually contains hundreds of web pages within it. Web users are only interested in navigating a site when it is properly organized, and the user locates information of his interest in several clicks. To do this, the structure of the websites requires their reorganization. This is achievable through ranking of the various web pages of website and thereafter rearrange web pages in their ranking. The most crucial or superior rank pages will be less clicks away in the homer page.

There are URLs of each web page. To address this issue with the help of the combinatorial optimization, URLs have received individual positive integer identifiers, and they appear in Table 1.

Table 1: Mapping of URLs with unique Ids

Webpages	WebP-1	WebP-2	WebP-3	….	WebP-n
Ids	1	2	3	….	n

As n is large may be hundreds or thousands and so on. Among them, we need to identify highest t (could be 10 or more) web pages.So there are very large number of possible combinations, equation 1, and it is computationally infeasible. Therefore, there is necessity to apply appropriate optimization algorithms and seek the best solution to this.

$$PossibleCandidateSolutionsi.e.SizeofSearchSpace = C_t^n = \frac{n!}{t! \times (n-t)!} \quad (1)$$

Where n is possible URLs and is large
t is top ranked 't' web documents

Metaheuristic optimization algorithms (MOA) require a starting population known as candidate solution. Table 2 shows a sample topmost web pages or candidate solution of 5.

Table 2: Sample topmost '5' web pages or candidate solution

Sample One Candidate Solution of Top-5 Web Pages				
Unique ids representing URLs generated randomly, Random(1,n)				
45	32	97	147	2

3.2 Criteria for Webpage Ranking

Good restructuring of a website structure needs proper assessment of relative significance of its webpages. In this, several requirements are taken into account that are based on user activities, relevance of content and structural importance of each page. The standards assist in both quantitative and qualitative evaluation of the utility of the webpages to allow a more informed and user-centered ranking system. The parameters chosen will help to be sure that the most valuable pages will be prioritized throughout the reorganization process. These multi-criteria are contained in the cost function that will be discussed in the next subsection.

3.2.1 Cost Function

The algorithms based on population require the evaluation of each candidate of the population at each iteration. The cost function is very critical in the success of any optimization algorithm. The higher the quality of cost, the more accurate the findings of algorithm will be. In the case of us, the cost function is dependent on several parameters as explained in Table 3.

Table 3: Fitness Function parameters

Parameters	**Abbreviation**
Cost of Keywords	CK
Content Relevance Score	CRS
Click-Through Rate	CTR
Cost of User's Engagement	CUE
Social Impact Score	SIS
PageRank	PR
Cost of Backlink Authority	CBA
Access Frequency	AF
Number of Distinct Visitors	DV
Hubs Score	HS
Authority Score	AS

Fitness function parameters are described next after that Cost function is formulated.

3.2.1 Cost of Keywords (CK)

Suppose that there are n keywords (k1, k2, k3, and so on kn) in query of the user. And web pages (WP1, WP2, WP3, etc., WPlmpm) in the site. The frequency in the keywords of WP1 f11, f12, f13, f14, f15, f16, f17, f18, f19, f1A, f1B, f1C, f1D in the frequency of WP1 of f11, f12, f13, f14, f15, f16, f17, f18, f19, f1A, f1B, f1C, f1D and the frequency in the keywords of WP2 f2

Therefore the Equation 2, price of chromosome based on keywords.

$$CK(WP_i) = \sum_{i=1}^{n} \quad \sum_{j=1}^{m} \quad f_{ij} \tag{2}$$

3.2.2 Content Relevance Score (CRS)

The Equation 3 finds that how relevant the content is to a query.

$$CRS(WP_i) = \frac{Number\ of\ Keywords\ Matches}{Total\ number\ of\ keywords} = \frac{p,(where\ p<n)}{n} \tag{3}$$

3.2.3 Click-Through Rate (CTR)

Click per second (WPi) of the user on some page on the site in the event that it is listed in the search results of a search engine. In this case we took into account the results of the search at Google shown on the Equation 4.

$$CTR(WP_i) = \frac{Number\ of\ times\ page\ WP_i\ clicked}{Number\ of\ Times\ WP_i\ appeared\ in\ SEO\ Result} \times 100 \quad (4)$$

3.2.4 Cost of User's Engagement (CUE)

Time duration on a particular web page is an important metric that reports the average amount of time spent by an individual on any site or web page. This provides the solution to user is searching behavior which in turn depicts the relevancy of web page. With the help of time duration of visited web page we compute its relevancy. If a web user do not find page as per his searching need he will not spent more time on it there may be case that he may remain idle for some time. In the table 3 shown below is users time remained in second by the user relating to specific URL hown in Table 4.

Table 4. Time Stayed for different Web page URL's

WebUser	Web_Url	Stayed Time
WebUser1	201.58.170.90	200
WebUser2	201.58.170.51	120
WebUser3	10.21.31.17	65
WebUser4	10.0.0.11	48
…	..	…
WebUser N	10.21.50.72	90

The Equation 5 captures the user's engagement with a page WP_i

$$CUE(WP_i) = \alpha.Time\ spent\ on\ page\ WP_i + \beta.(1 - Bounce\ Rate) \quad (5)$$

Where time spent on the page by user is the average time of stay WPi on a page.
Bounce Rate: Percentage of one page sessions.
Each of the metrics has weights that are assigned to it according to its importance as α and 8.

3.2.5 Social Impact Score (SIS)

SIS gives the influence of social participation of a web page. The Equation 6 gives the amount of engagement that a web page gets in the social media sites. It shows the effectiveness of the audiences in engaging with the content and it is usually a part of the online presence and impact to the site.

$$SIS(WP_i) = \alpha.Shares\ on\ WP_i + \beta.Comments + \gamma.Likes \quad (6)$$

Where α, β, and γ are weights assigned to each type of interaction.

3.2.6 PageRank (PR)

The Equation 7 PageRank measures the importance of a web page WP_i based on the quality and quantity of backlinks.

$$PR(WP_i) = (1 - d) + d.(\sum_{i=1}^{\varphi} \quad \frac{PR(LP_i)}{C(LP_i)} \quad (7)$$

Where;

- d is damping factor
- $PR(LP_i)$ is page rank of linking page

- $C(LP_i)$ is number of outlinks on page LP_i
- φ is number of pages linked to WP_i

3.2.7 Cost of Backlink Authority (CBA)

The quality and general power of the back links to a webpage is determined using cost of BA. The Equation 8 is also applied in SEOs to establish the extent of influence of a backlink on a web page ranking.

$$CBA(WP_i) = \sum_{i=1}^{l} \frac{DA(WP_i) \times R(WP_i)}{C(WP_i)} \quad (8)$$

Where;

- $DA(WP_i)$ is domain authority of linking page WP_i
- $R(WP_i)$ is relevance score of target page WP_i
- $C(WP_i)$ is number of outbound links on WP_i
- l is total number of backlinks

3.2.8 Access Frequency (AF)

Access frequency of webpage the total number of visits of each user at a given webpage. This indicator allows us to assess the extent to which there are new users on the site the relevancy is more when web page is visited more. Thus the check of access recurrence to webpage is use related measure. In table 5 below the access count on different URL regardless of Id are indicated.

Table 5. Acces Count for different Web page URL's

WebUser	Web_Url	Access Count
WebUser1	201.58.170.90	35
WebUser2	201.58.170.51	22
WebUser3	10.21.31.17	26
WebUser4	10.0.0.11	49
..	..	..
WebUser N	10.21.50.72	28

$$AF(WP_i) = \frac{Total\ Number\ of\ Visits\ of\ WP_i}{Time\ Period} \quad (9)$$

Where;

- Total Number of Visits: The total number of visits to the webpage, including both unique and repeat visits to the webpage within a given period is Shown in The Equation 9.
- The Equation 10 shows Time Period: This is the time that the visits will be assessed and the time average is usually in days.

$$Weighted\ Access\ Frequency(WP_i) = \frac{\sum_{i=1}^{U} (WP_i \times Engagement\ Score)}{Time\ Period} \quad (10)$$

Where, Engagement Score: A metric that shows user behavior either through clicks and time on the site, bouncing rate or conversion.

3.2.9 Number of Distinct Visitors (DV)

The role of unique visitor on web page is very important in the process of establishing the relevance of web page. Unique Visitors: The Equation 11 shows the number of traffic that came to visit the web page during the period of time being measured. When a person visits the web page many times within the period of time, he/she is only counted once. Considering the server log, this data can be based on it. A sample data is presented in table 6.

Table 6. UniqueVIsitors for different Web page URL's

UniqUser	Web_Url	Unique Visitor
WebUser1	201.58.170.90	121
WebUser2	201.58.170.51	26
WebUser3	10.21.31.17	85
WebUser4	10.0.0.11	48
…	..	…
WebUser N	10.21.50.72	67

$$DV(WP_i) = \sum_{j=1}^{U} \quad count(Visitor_j(WP_i)) \tag{11}$$

3.2.10 Hubs Score (HS)

Hub is a web page, which links to numerous other good content web pages on a particular subject known as authoritative page whereas an authority is a webpage that is thought of as reliable source of information and which is referenced by numerous different hubs as indicated in Figure. 2.

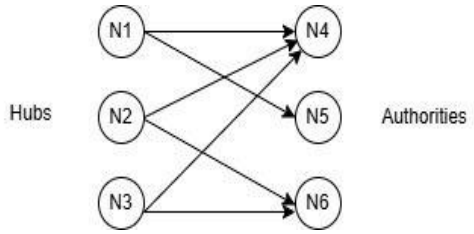


Figure.2 Hubs and Authorities

$$Hub(WP_i) = \sum_{j=1}^{|q|} \quad Auth(q(j)) \tag{12}$$

Where, q=set of pages linked to by page p (its outlinks) is shown in The Equation 12

$$q = \{AP_1, AP_2, \dots, AP_\omega\}, \omega = number\ of\ outlinks$$

3.2.11 Authority Score (AS)

Authority score of a page WP_i is calculated similarlyusing The Equation 13.

$$Auth(WP_i) = \sum_{j=1}^{|p|=\varphi} \quad Hub(p(j)) \tag{13}$$

Where, q=set of pages linked to by page p (its outlinks)

$$p = \{H_1, H_2, \dots, H_\varphi\}, \varphi = number\ of\ outlinks$$

Finally, Cost function can be defined as in The Equation 14

$$Cost(WP_i) = d1.CTR(WP_i) + d2.CUE(WP_i) + d3.SIS(WP_i) + d4.PR(WP_i) + d5.CBA(WP_i) + d6.Weighted\ Access\ Frequency(WP_i) + d7.DV(WP_i) + d8.Hub(WP_i) + d9.Auth(WP_i) \quad (14)$$

where d1, d2, d3, d4, d5, d6, d7, d8, and d9 are normalization constants.

3.3 Working of MA-BBO-PSO Hybrid

To restructure the websites, the suggested WDR_MA_BBO_PSO hybrid metaheuristic algorithm will be used as a combination of three optimization algorithms, namely, MA, BBO, and PSO. The algorithm is developed to successively augment the candidate solutions through a tiered optimization strategy where each element has its distinct advantages in exploration and exploitation of the solution space. First, a wide range of possible solutions (considering ranked website structures) is created. These solutions are then refined by MA. Then, BBO is performed by migration and mutation operations which promote information exchange between solutions and preserve diversity. Lastly, PSO attempts to locate the best solutions of the world and accelerate the convergence. Combinely MA, BBO and PSO facilitate the algorithm in escaping to the stuck in local maxima and in preserving the diversity of exploration. PSO enables fater convergenc to receive an optimal list of highest ranked webpages that would be applied to restructure the web site. This will guarantee that the most pertinent pages are made near to the homepage thus improving the efficiency of navigation in the websites.

Overall Flow of WDR_MA_BBO_PSO:

1. *MA Phase:*
   - Generate diverse population
   - Apply local refinement using Iterative Improvement or Simulated Annealing
2. **BBO Phase:**
   - Use MA-refined population
   - Apply migration and mutation for further optimization
3. **PSO Phase:**
   - Initialize particles from BBO-refined solutions
   - Perform velocity and position updates
   - Return globally optimal webpage rankings

3.3.1. Initialization using Memetic Algorithm (MA)

Memetic algorithm generates initial population, the process begins with this. MA has been chosen because it can combine global searching using evolutionary operators and also local optimization using problem specific heuristics which in our case is the rank of set of webpages. It creates a diversed original population in terms of such characteristics as Rank or cost of one candidate solution is calculated based on Cost of Keywords, Content Relevance Score, Click-Through Rate, Cost of User Engagement, Social Impact Score, PageRank, Cost of Backlink Authority, Access Frequency, Number of Distinct Visitors, Hubs Score, Authority Score. Once the local improvement is done by the means of iterative improvement, the refined population is subjected to the next step.

3.3.2 Iterative Improvement

MA utilizes the local search method, Iterative Improvement. It starts with the initial population of candidate solutions and then continues to refine the solutions on the basis of the cost function. The process is repeated until a set number of iterations (MAX_ITERation) have been completed through the separation and execution of the algorithm and monitoring of a decrease in the cost.

Algorithm 1: Iterative Improvement Local Search
Procedure: IterativeImprovement(initialSolution)

1. *begin*
2. *Set current_Solution ← initial_Solution*
3. *Set iteration ← 1*
4. *while (iteration ≤ MAX_ITERATION OR no significant improvement in currentSolution) do*
   a. *Evaluate the current_Solution*
   b. *candidate_Solution ← generateCandidateSolution(current_Solution)*
   c. *if isBetter(candidate_Solution, current_Solution) then*
      i. *current_Solution ← candidate_Solution*
   end if
   d. *iteration ← iteration + 1*
   end while
5. *return current_Solution*
6. *end*

3.3.3 Simulated Annealing

In the MA section of the proposed hybrid model, the Simulated Annealing (SA) is employed. It can prevent the local maxima, particularly when the applicants get stagnant in the search process. SA presents a controlled random procedure that sometimes takes inferior solutions. The probability of such acceptance decreases with the decrease in the temperature(cost) in the following iteration. The method approximates the procedure of heating a substance to a elevated temperature and subsequently, gradually, lowering the temperature to a low energy state. This simulated cooling process is done by slowly reducing the rate of cooling and this increases the probability of arriving at a globally optimal solution. Computationally, this can be modeled by SA creating a new candidate state of the current state by generating a random displacement state. This new state is in turn examined.

- If better solution it is accepted but sometimes poor solution is also accepted with certain probability calculated using equation (15).

$$P = e^{-(E_new_state - E_current_state)/T} \quad (15)$$

Here, E_new_state and E_current_state represent the cost (or energy) of the new and current states, respectively, and T is the current temperature.

Algorithm 2: Simulated_Annealing()

1. Initialize:
a. Set current_solution S ← RandomState()
b. Set best_solution maxS ← S
c. Set initial_temperature, T ← T_initial
d. Define minimum_temperature T_min
e. Set other algorithm parameters (e.g., cooling rate, iterations per temperature)

2. While T > T_min do
a. For i = 1 to Iterations Per Temperature do:
i. Generate a candidate_solution S' ← Neighbor(S)
ii. Evaluate cost(S) and cost(S')
iii. If cost(S') < cost(S) then:
• S ← S'
• If cost(S') < cost(maxS) then:
maxS ← S'
iv. Else:
• ΔE ← cost(S') - cost(S)
• p ← exp(-ΔE / T)
• If random(0,1) < p then:
S ← S'
b. Update temperature: T ← α × T (cooling schedule)
3. Return maxS // Best solution found
End Procedure

3.3.4 Memetic Algorithm

The Memetic Algorithm (MA) is a population-based metaheuristic algorithm, which aims at searching optimal solutions to a large search space, like a set of web documents. It integrates the global search capability of evolutionary algorithms and local optimization capability of heuristic search methods that give it even faster and more precise convergence than traditional Genetic Algorithms (GAs). In MA, every candidate solution (chromosome) undergoes genetic operations such as crossover and mutation and then local search is done to enhance quality of candidate solutions. When it comes to the ranking of web pages and restructuring of websites, MA finds its use in optimizing the process of organizing web pages in huge and dynamic data sets.

Algorithm 3: Memetic Algorithm for Web Document Optimization

Input: *Initial web document population, parameters (population size* `pop_size`*, crossover rate* `cross_rate`*, mutation rate* `mutate_rate`*, max generations* `max_gen`*)*
Output: *Optimized set of ranked document*

1. ***Begin***
2. *Encode initial solutions into chromosome format*
3. *Initialize population with* `pop_size` *individuals*

4. Set parameters: `max_gen`, `cross_rate`, `mutate_rate`
5. Set `gen ← 0`
6. ***While*** *(*`gen < max_gen`*)* ***do***
7. Evaluate fitness of all individuals in the population
8. Apply evolutionary operators:
9. a. Selection
10. b. Crossover (with probability `cross_rate`*)*
11. c. Mutation (with probability `mutate_rate`*)*
12. Generate new offspring population
13. ***For each*** *offspring* ***do***
14. Apply local search (e.g., Simulated Annealing or Hill Climbing) for intensification
15. ***EndFor***
16. Update current population with new offspring
17. `gen ← gen + 1`
18. ***EndWhile***
19. Apply final local search to the best individual
20. ***Return*** *optimized solution*
21. ***End***

3.3.5 Biography Based Optimization

Biogeography-Based Optimization (BBO) is a metaheuristic based on population population inspired by geographical distribution of species. In BBO candidate solutions (habitats) exchange features by immigration and emigration. It has proven to be useful in different areas including feature selection, and classification as well since it has balance of exploration and exploitation. BBO can be used in the context of reorganization of web pages to refine the ranking of web pages by utilizing migration and mutation operators to preserve diversity and enhance convergence.

Migration Operator:

With current habitat Hi species Si accommodates. Sm is the maximum number of species, C i and β i are immigration rate and emigration rate of Hi, respectively, which are modified as Equation 16 and 17.

$$C_i = T * (1 - \frac{S_P}{S_m}) \quad (16)$$

$$\beta = e * \frac{S_P}{S_m} \quad (17)$$

Mutation operator:

The Equation 18 shows Mutation rate M_i and species probability P_i. Species probability P_i is decided by number of species S_i [1] ($S_{max} = N$ and $S_i = S_{max} - i$).

$$M_i = M_{max} * (1 - \frac{P_i}{P_{max}}) \quad (18)$$

where M_{max} is the maximum mutation rate and P_{max} is maximum species probability.

Algorithm 4: -Based Optimization (BBO)

Input: Objective function f(x), population size N, maximum generations G,
maximum mutation rate M_max, migration parameters T, e,
species count range [0, S_max]

Output: Best solution found
1. Initialize population of candidate solutions (habitats) H = {H1, H2, ..., HN}
2. Evaluate fitness f(Hi) for all habitats in H
3. Set generation g = 0

4. while g < G do
5. Compute immigration rate λ_i and emigration rate μ_i for each habitat Hi:
$\lambda_i = T \times (1 - Si / S_{max})$
$\mu_i = e \times (Si / S_{max})$
6. For each habitat Hi in H:
7. Select Hj (j ≠ i) based on μ_j
8. Migrate features from Hj to Hi with probability λ_i
9. Compute species probability P_i for each Hi
10. Apply mutation with rate:
$M_i = M_{max} \times (1 - P_i / P_{max})$
11. Evaluate new fitness f(Hi) for all modified habitats
12. Update the population by retaining best solutions
13. g = g + 1
14. end while
15. Return the habitat with the best fitness

3.3.5 Particle Swarm Optimization

One of the original and most commonly used swarm intelligence algorithms is describing the Particle Swarm Optimization (PSO) which is based on the collective behavior of bird flocks or fish schools. The basic idea is that every creature (this will be referred to as a particle) is a possible solution which is traversing the search space and experiences a learning process based on its own experience and the knowledge of the overall swarm. This cooperation assists the swarm to optimize or approach optimal solutions.

In PSO, every particle modulates its path relying on its individual best position (pBest) and the best-known universal position which was discovered by the swarm (gBest). The active redistribution of the positions and velocities of the particles enables the swarm to manage the exploration (global search) and exploitation (local refinement) which is essential in preventing local optima and achieving convergence in complex solution ontologies like web navigation optimization.

Mathematical Formulation

Each particle i has:

- A position vector Xi(x),
- A velocity vector Vi(x),

Where pBesti is a personal best position.

- A globally most famous post gBest.

The velocity and position of each particle are updated iteratively as follows The Equation 19 &20:

$$V_i(x+1) = p \times V(x) + s1.r_1(pBest_i - X_i(x)) + s2.r_2(gBest - X_i(x) \quad (19)$$

$$X_i(x+1) = V_i(x) + V_i(x+1) \quad (20)$$

Where:

and p is the inertia weight of influence at the moment of previous velocities,

- s1 and s2 are cognitive and social (acceleration) factors,
- r1, r2 are random numbers in [0,1],
- Vmax- This limits the highest speed.

This model allows the particles continue to alternate the search direction based on the personal learning and the effect of the group and is also relevant to dynamic and high-dimensional problems which is the clustering and reorganization of web pages.

- p is the null weight that regulates the effects of past velocity,
- s1 and s2 are factors of acceleration (cognitive and social factors)
- r1, r2 are random numbers in [0,1],
- Vmax is used to limit the top speed.

This model enables the particles to constantly regulate their search pathway with respect to personal learning as well as group impact and thus is adequately suitable to dynamic and high-dimensional issues such as webpage clustering and reorganizatio

Algorithm 5: PSO Algorithm

 # Parameters:
 # d: number of dimensions, N: number of particles, max_iter: maximum iterations

 # w: inertia weight, c1: cognitive coefficient, c2: social coefficient

1. Initialize population of N particles with random positions and velocities.

2. For each particle i, evaluate fitness and store pBest.

3. Identify global gBest among all particles.

4. For each iteration until max_iter:

a. For each particle i:
i. Update velocity:
*V_i = w * V_i + c1 * r1 * (pBest_i - X_i) + c2 * r2 * (gBest - X_i)*
ii. Clip velocity if out of bounds.
iii. Update position:
X_i = X_i + V_i
iv. Evaluate fitness:
Update pBest_i and gBest if needed.

5. Return the global best particle as the final solution.

3.4 Proposed Hybrid MA-BBO-PSO Optimization

The Figure 3 shows the Proposed Hybrid MA-BBO-PSO Optimization

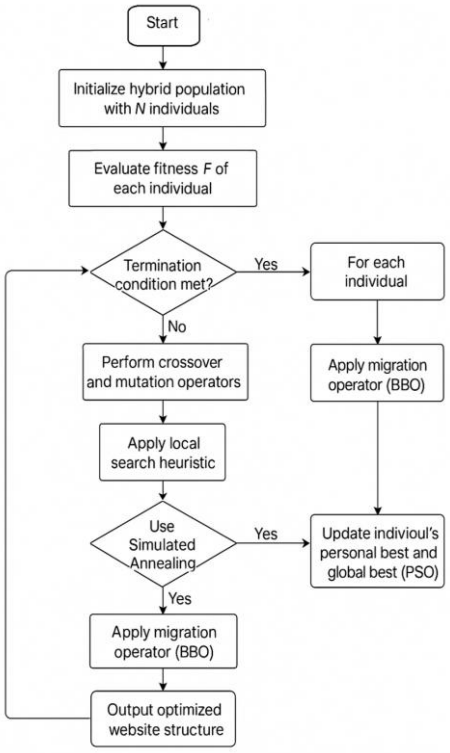


Figure 3: Proposed Hybrid MA-BBO-PSO Optimization

Algorithm 6: Hybrid MA-BBO-PSO for Webpage Ranking and Website Reorganization

***Input**:*

- *Website structure graph G=(V,E)G = (V, E)G=(V,E),*
- *User navigation data,*
- *Fitness function F (e.g., minimizing average click distance, improving visit coherence),*
- *Parameters for MA, BBO, and PSO.*

***Output**:*

- *Optimized website link structure and ranked webpages.*

Begin

1. ***Initialize** the hybrid population with N individuals (solutions), each representing a candidate website structure.*
2. ***Evaluate** fitness F of each individual.*
3. ***Repeat until termination condition is met (e.g., max iterations or convergence):***
4. ***For** each individual in the population **do***
5. *Apply **Memetic Algorithm (MA) local search**:*
   a. *Apply crossover and mutation operators to create offspring.*
   b. *Perform **local refinement** using a problem-specific heuristic.*
   c. *Apply **Simulated Annealing** if solution stagnates (escape local optima).*
      - Accept worse solution with probability $P=e^{-\Delta E/T}$, where ΔE is fitness difference.
   d. *Update solution with improved version.*

6. ***Apply Biogeography-Based Optimization (BBO):***
   a. *Compute immigration (λ_i) and emigration (μ_i) rates using:*

$$\lambda_i = \lambda_{max} * (1 - \frac{S_i}{S_{max}})$$

$$\mu_i = e.\frac{S_i}{S_{max}}$$

   b. *Migrate features between individuals probabilistically.*
   c. *Apply mutation:*

$$M_i = M_{max} * (1 - \frac{P_i}{P_{max}})$$

   d. *Update species probability for each habitat.*

7. ***Apply Particle Swarm Optimization (PSO):***
   a. *For each particle (solution), update velocity using:*

$$V_i(x+1) = p \times V(x) + s1.r_1(pBest_i - X_i(x)) + s2.r_2(gBest - X_i(x)$$

   b. *Update position:*

$$X_i(x+1) = X_i(x) + V_i(x+1)$$

   c. *Clip velocity within bounds.*
   d. *Evaluate fitness and update pBestpBestpBest and gBestgBestgBest.*

8. ***Elitism****: Retain top-performing individuals for the next generation.*
9. ***End Repeat***
10. ***Return*** *the best-ranked and reorganized website structure.*

End

4 Result and Discussion

4.1 Experimental Setup

All the algorithms were written in Java (JDK) and run in a windows platform. Synthetic datasets were created to reproduce the properties of the web usage, the content and the hyperlink structure in order to create the simulation of real-world scenarios. Table 6 presents the statistical analysis of the dataset that includes major characteristics of the dataset frequency of keywords, frequency of access, number of visitors is shown in the Table 7, the duration of session, and properties of hyperlinks (hubs and authorities).

Table 7. Statistical Analysis of Dataset

	page_id	Keyword_fre	Access_fre	Num_of_visitors	Tot_duration	Num_of_hubs	Num_of_auths
count	20000	20000	20000	20000	20000	20000	20000
mean	9999.5	49.76	49.8551	49.31115	49.6061	49.53955	49.4229
std	5773.64703	29.125765	28.9932826	28.825958	28.8384765	30.2737256	30.2198514
min	0	0	0	0	0	0	0
25%	4999.75	25	25	24	25	26	21
50%	9999.5	50	50	50	49	47	47
75%	14999.25	75	76	74	75	73	73
max	19999	100	100	100	100	100	100

4.2 GA Parameter Tuning and Selection of Top-K Web Documents

The Genetic Algorithm (GA) was initially used to determine the top-K ranked web documents. The tuning of the parameters was done by changing the population size between 50 and 300

and by using different crossover and mutation rate configurations. The configurations of the parameters are summarized in table 8.

Table 8. Specification of Genetic Algorithm Parameter

Parameter	Value
Initial Population Size	50,100, 150,200, 250,300
Crossover rate	50%, 60%, 70%, 80%, 90%
Mutation rate	0.05, 0,06, 0.07, 0.08, 0.09, 0.10
Size of Chromosome(Web Document)	5, 6, 7, 8, 9, 10
Selection method (selection of population for crossover	Binary tournament selection (BTS)
Crossover type	CyclicCrossOver

Table 9 shows the performance of GA according to the effect of varying mutation rates (population size =50 and Top-K =5). The findings suggest that the mutation rate with a 10 percent is optimum in terms of fitness.

Table 9. Mutation Rate vs Fitness Quality

Finding most suitable crossover rate and Mutation rate						
Popsize=50, SM=BTS, Order one crossover, Top-K=5						
CR	MR					
	5%	**6%**	**7%**	**8%**	**9%**	**10%**
	Quality					
50%	2182	2181	2223	2239	2214	2347
60%	2361	2348	2352	2344	2143	2143
70%	2170	2293	2181	2062	2086	2376
80%	2171	2278	2267	2238	2300	2205
90%	2126	2243	2298	2333	2336	2351
Average	2202	2268.6	2264.2	2243.2	2215.8	2284.4

Virtual cross over rates were then assessed (fixing mutation at 10 percent). Table 10 shows that a 70 percent crossover rate is always able to produce better results when compared to different values of Top-K.

Table 10. Crossover Rate vs Fitness Score

Fitness score for TOP-K results					
Popsize=50, SM=BTS, Order one crossover, MR=10%					
TOP-K	CR				
	50%	**60%**	**70%**	**80%**	**90%**
	Quality				
5	2299	2338	2268	2437	2341
6	2556	2827	2924	2754	2863
7	3269	3110	3272	3140	3256
8	3428	3712	3613	3758	3615
9	4080	4019	4301	4051	3822
10	4429	4560	4421	4692	4550

These optimal parameter values (CR = 70%, MR = 10%) were adopted for all further GA, MA, and hybrid model experiments.

4.3 Comparative Analysis with Other Metaheuristic Algorithms

An in-depth performance analysis of the proposed hybrid algorithm (WDR_MA_BBO_PSO) and the other benchmark algorithms such as, WDPII, WDPSA, WDPGA, WDPMA, WDPBBO and WDPPSO were done. Table 11 gives the parameter settings of each algorithm.

Table 11. Algorithmic Parameter Settings

WDPII	Population Size=300, Number of neighbors=10
WDPSA	Population Size=300, Initial Temperature T = RankValue(intital set of top-k docs), Initial Optimal State=Randomly selected top-k ranked web documents, Neighbors=10, T=0.9*T
WDPGA	Population Size=300, CR=70%, MR=0.10, Selection=Binary Tournament
WDPMA	Population Size=300, CR=70%, MR=0.10, Neighbors=10, Selection=Binary Tournament
WDPBBO	Population Size=300, Number of habitats=5, Migration rate=30%
WDPPSO	Population Size=300, C1 = 1, C2 = 1 and W = 1
WDR_MA_BBO_PSO	Population Size=300 and above settings

The algorithms were tested with Top- K = 10 at different population sizes (50 to 300). As indicated in Table 12, the proposed hybrid strategy is always better in terms of fitness score as compared to other algorithms.

Table 12. Fitness Score vs Population Size (Top-K = 10)

Fitness score for top-K results for different population sizes							
POPULATION ↓	**WDPII**	**WDPSA**	**WDPGA**	**WDPMA**	**WDPBBO**	**WDPPSO**	**WDR_MA_BBO_PSO**
50	2286	2862	3171	3643	3903	4506	4603
100	2325	2771	3417	3665	4331	4608	4734
150	2350	2908	3397	3872	4406	4772	4789
200	2430	2981	3447	3887	4308	4805	4923
250	2414	2948	3405	3841	4293	4823	4974
300	2413	2885	3412	3827	4291	4836	5045

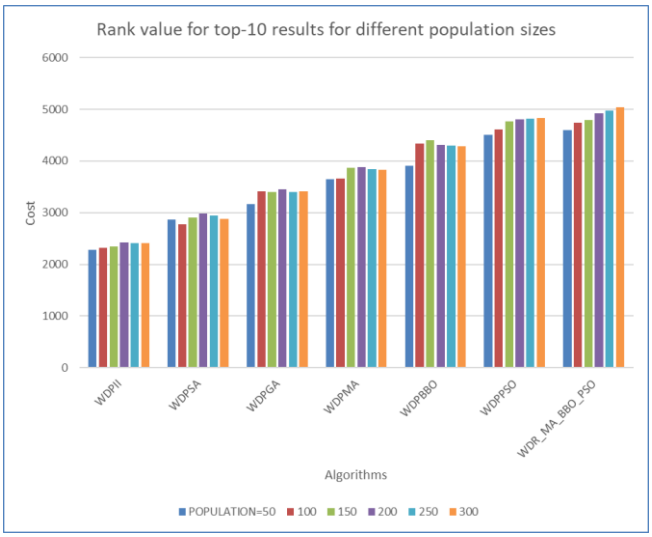


Figure. 4. Rank value for top=10 results for different population sizes.

This is also compared with Table 13 which shows the performance of all the algorithms on Top-K 5-10. The proposed technique has the highest fitness in all instances, showing its strength in the process of picking the good quality web documents is shown in Figure 4.

Table 13. Top-K Selection vs Cost of Selection

Top-T	WDRII	WDRSA	WDRGA	WDRMA	WDRBBO	WDRPSO	WDR_MA_BBO_PSO
5	2598	3090	3584	4075	4567	5098	5258
6	2599	3095	3622	4102	4610	5074	5315
7	2616	3127	3630	4110	4604	5084	5289
8	2604	3112	3616	4169	4605	5099	5122
9	2487	2975	3487	4000	4428	4926	5099
10	2511	3015	3468	4015	4444	4947	5047

4.4 Generation-wise and Top-K Analysis

Additionally, the algorithms, WDPII, WDPSA, WDPGA, WDPMA, WDPBBO and WDPPSO and WDR_MA_BBO_PSO have been compared generation wise and the results are depicted in Figure. 5. In this case, too, the proposed algorithm can choose of higher quality Top-T webpages.

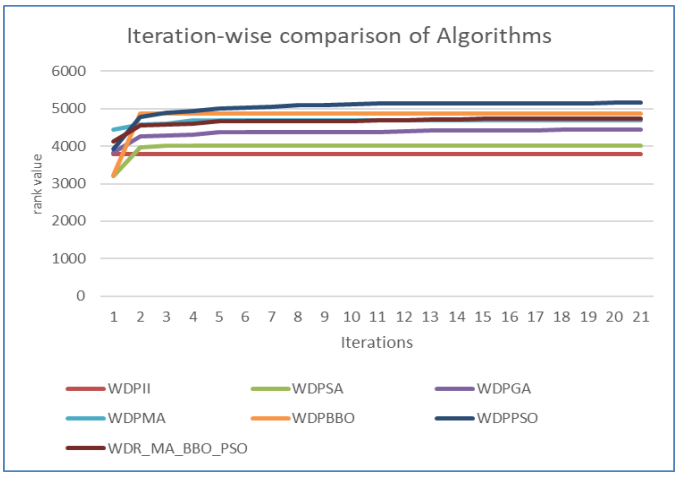
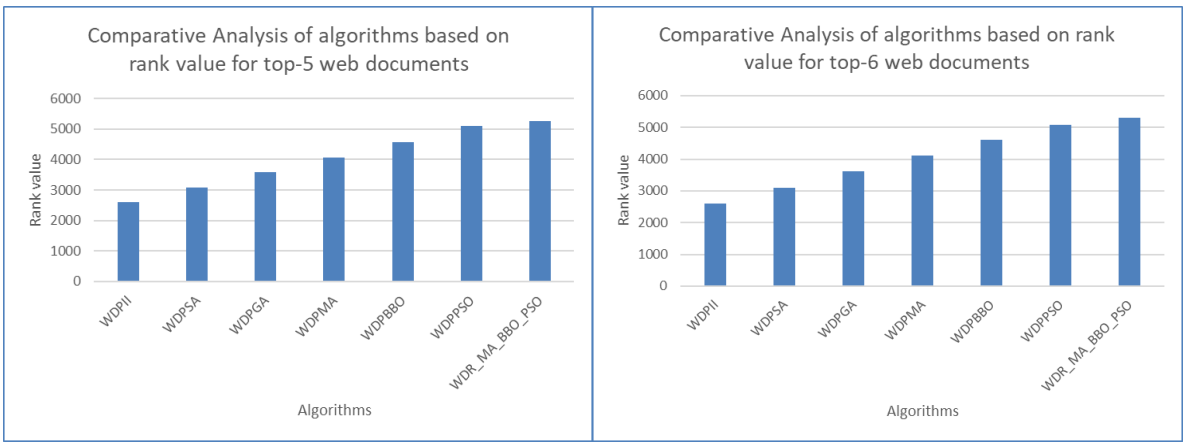


Figure. 5. Generation-wise and Top-K Analysis

The trends of convergence on a generation basis illustrated in Figure 6 indicate that the convergence of WDR_MA_BBO_PSO is more stable and rapid. The performance of Top-K (K = 510) is compared in detail based on a population size = 300 as shown in Figure 5. In all K-values, the algorithm proposed consistently beats the rest of the techniques, and the quality of the solutions was increased on average by 2.89 percent.



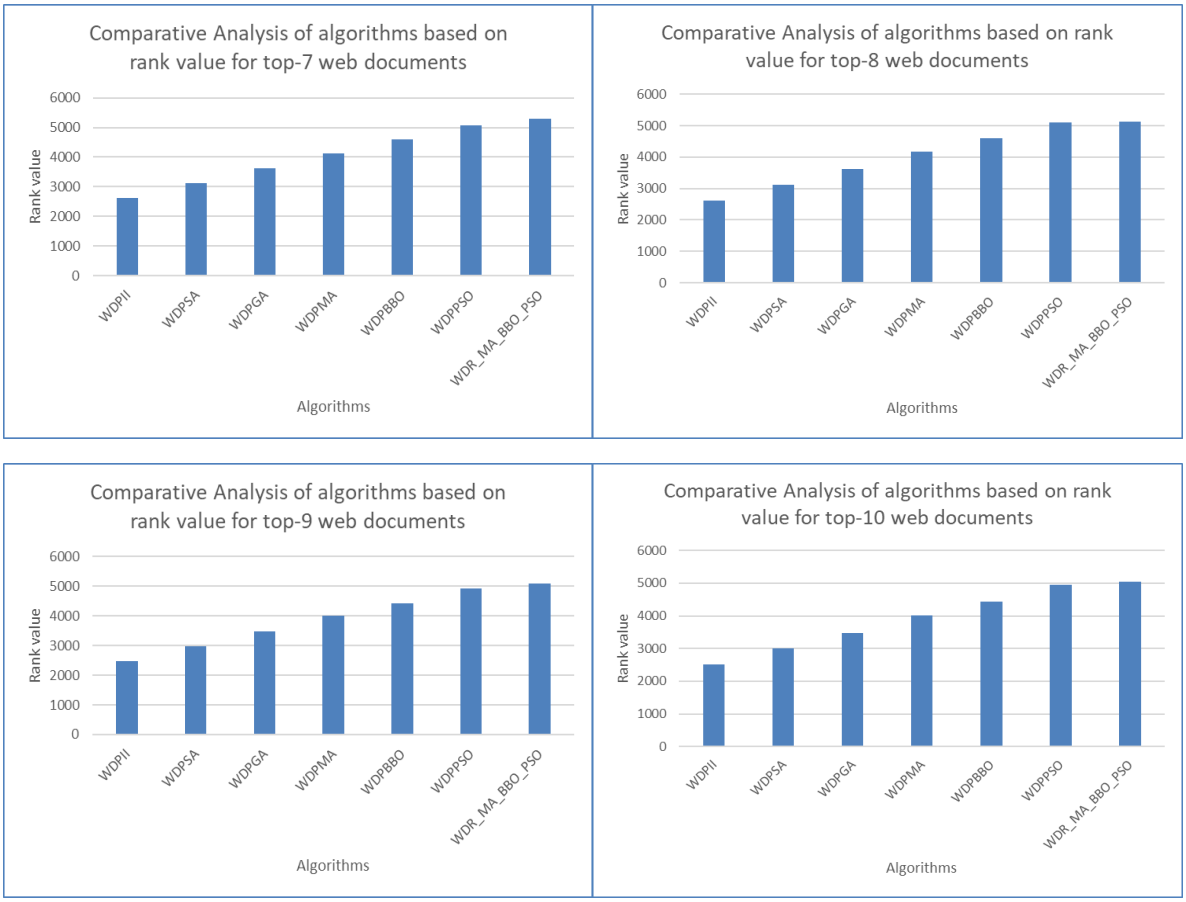


Figure. 6. Comparative analysis of algorithms

4.5 Evaluation on Real Traffic and Log Datasets

The suggested approach was also tested on a real traffic dataset which comprised the features like page views, session duration, bounce rate and conversion rate. The user navigation sequences and the request metadata were also extracted using server logs. Table 14 shows the ranking accuracy metrics (Precision, Recall, F1 Score) values.

Table 14. Ranking Accuracy

Algorithm	Precision (%)	Recall (%)	F1 Score
PageRank	85	70	0.76
HITS	82	75	0.78
BM25	90	65	0.76
TF-IDF	80	68	0.74
RankBrain	88	80	0.84
WDPIIGA	92	79	0.84

The hybrid approach has a Precision of 92 and Recall of 79, which is better than the page rank, the HITS and RankBrain baselines.

Regarding user engagement indicators, Click through rate (CTR) and average time, the Table 15 proves that the proposed algorithm performs the best, with CTR of 39% and average time of 142 seconds.

Table 15. User Engagement

Algorithm	Average CTR (%)	Average Time Spent (seconds)
PageRank	30	120
HITS	28	115
BM25	32	110
TF-IDF	26	100
RankBrain	35	140
WDR_MA_BBO_PSO	39	142

Table 16 gives the ranking stability in terms of the mean rank position and standard deviation of the ranking on multiple queries. WDR_MA_BBO_PSO method is stable in terms of competition with an average rank of 6.7 and SD of 1.3.

Table 16. Ranking Stability

Algorithm	Average Rank Position	Standard Deviation
PageRank	5.4	1.2
HITS	6.1	1.5
BM25	4.8	0.9
TF-IDF	7.3	1.8
RankBrain	5	1
WDR_MA_BBO_PSO	6.7	1.3

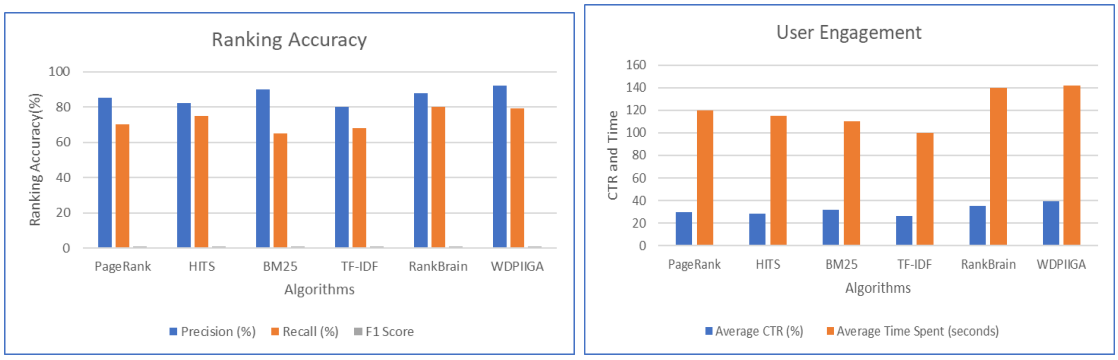


Figure. 7: Comparison of Raking accuracy and User engagement

Based on Figure. 7 it will be possible to admit that the proposed WDR_MA_BBO_PSO will be effective in terms of the accuracy of the offered ranking and user engagement indicators, which proves its efficiency in offering the relevant result.

4.6 Execution Time and Computational Efficiency

In order to analyse the efficiency of the execution, consider reporting or estimating:

1. Run Time (seconds or milliseconds) of various population sizes (e.g. 100, 200, 300).
2. Convergence Speed: Generations to optimal or near optimal fitness.
3. Computational Overhead: hybridization Time/memory trade-off.
4. Complexity at Time: The Big-O notation (qualitative).

5. Speedup Ratio: Speedup=Execution Time of baselineExecution Time of proposed= Speedup=Execution Time of proposedExecution Time of baseline are shown in the table 17

Table 17: Execution Time and Computational Efficiency

Algorithm	Execution Time (s)	Generations to Converge	Avg Fitness @ Gen 50	Speedup over GA
WDPII	12.4	75	4021	1×
WDPSA	13.8	88	4175	0.90×
WDPGA	10.9	68	4380	1.13×
WDPMA	11.5	64	4561	1.08×
WDPBBO	10.6	59	4620	1.17×
WDPPSO	9.4	53	4812	1.32×
WDR_MA_BBO_PSO	7.9	47	5045	1.57×

In order to assess the computational efficiency of the suggested hybrid algorithm WDR_MA_BBO_PSO, we performed the performance comparisons in relation to traditional metaheuristic methods (GA, SA, BBO, PSO, etc.) in terms of various population sizes. The mean execution time was measured in more than 30 independent executions. The proposed hybrid model was found to be more successful compared to others, as demonstrated in Table X, in the convergence speed and the runtime.

It is worth mentioning that, WDRMABOpsos had a convergence in only 47 generations, whereas GA and WDPII required 68 and 75 generations respectively, whereas it maintained high fitness scores. The mean time the suggested method took to execute was 7.9 seconds which is about 1.57 times less than the time needed by the basis GA method. This is attributed to the layered optimization strategy where MA works to provide a fast local optimization, BBO provides diversity due to migration and PSO works and speeds up convergence of global search.

This hybridization is an appropriate balance between exploration and exploitation and eliminates unnecessary computations and premature convergence and increases scalability and efficiency, particularly in large datasets.

4.7 Observations and Implications

The results of the experiment allow making a few interesting observations and practical conclusions:

- WDR_MA_BBO_PSO is a hybrid model, which is capable of combating global exploitation and local exploration sequentially by using Memetic Algorithm (MA), Biogeography-Based Optimization (BBO) and Particle Swarm Optimization (PSO). This cascade raises the efficiency of this algorithm in searching the search space and preventing premature convergence.
- The method exhibits quality webpage rankings in a wide range of Top-K document selections with high fitness scores, high convergence rates and high generalization in diverse population sizes and parameter settings.
- Outcomes of the proposed hybrid system Compared to the conventional ranking algorithms like PageRank, HITS, TF-IDF, and BM25, the offered hybrid approach generates much higher results in terms of ranking accuracy (precision: 92%, recall:

79%), and higher user engagement values such as click-through rate (CTR: 39%) and the average time spent (142 seconds).

- Also, the stability in the ranking (standard deviation: 1.3) of the model demonstrates its strength and dependability in dynamic and large-scale web setting.

These findings indicate that hybrid metaheuristic methods may be critical in intelligent web navigation, adaptive web site organization, as well as in individualized content delivery - aspects that are of importance in enhancing the performance of systems and the user experience in the contemporary information retrieval systems.

5 Conclusion

An effective webpage ranking and web site reorganization hybrid metaheuristic, WDR_MA_BBO_PSO, was suggested in this paper in terms of the multi-criteria evaluation. The algorithm combines local refinement capabilities of Memetic Algorithm (MA), Biogeography-Based Optimization (BBO) capabilities of sharing desirable features via migration, and Particle Swarm Optimization (PSO) capabilities of global convergence acceleration. On synthetic and real-world datasets of the web, the proposed hybrid approach was tested based on such features as keyword frequency, access patterns, session duration, and hub-authority scores. The proposed approach was superior to the use of baseline algorithms, such as GA, SA, BBO, PSO, and the classical models such as PageRank and HITS. Particularly, it got the best fitness score of 5045 when selecting the Top-10 documents (population size = 300) which was far much better than the other methods. The proposed model also achieved a high accuracy of 92 percent and recall of 79, which is considered highly relevant and covers all the relevant documents that were retrieved. To capture the attention of users, the model had an average of 39 percent of click through rate (CTR) and 142-second average time spent, which validated its capability of facilitating quality user clicks. In addition, the technique had good stability of ranking with standard deviation of 1.3 and average position of ranks being 6.7 on various queries. These findings confirm the suitability of the hybrid model to the intent of users and relevance of documents on a scale basis. The algorithm can be expanded into real-time accommodation of websites based on session recording, heatmap and user feedback in future work and be further combined with semantic analysis based on deep learning in order to support intelligent web navigation and personalization.

The algorithm can be further developed in the future to dynamically restructure web sites based on live user feedback, heatmap, or video recordings. Furthermore, the integration of deep learning models in semantic document understanding and integration with live recommender systems may help to provide the even greater practical implications of the study.

Declaration

Data Availability: No dataset was generated or analyzed in this study

Conflicts of Interest: The authors declare that they have no conflicts of interest regarding the publication of this manuscript.

Funding Statement: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution: *Ravi Kumar:* Conceptualization, methodology, algorithm design, software implementation, experimentation, data analysis, validation, and manuscript writing. *Santosh Kumar:* Supervision, methodology refinement, review and editing, validation, and overall guidance throughout the research.

Ethical Approval: This study did not involve human participants, human tissue, animals, or any interventions requiring ethical approval. Therefore, ethical approval was not required.

Consent to Participate: Not applicable. This study did not involve human participants.

Consent for Publication: Not applicable.

Competing Interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Code Availability: The implementation of the proposed **WDR_MA_BBO_PSO** algorithm is available from the corresponding author upon reasonable request.

Acknowledgement: The authors would like to thank their respective institutions, ABES Engineering College, Ghaziabad, and Galgotias University, Greater Noida, for providing the necessary facilities and academic support to carry out this research. The authors also appreciate the valuable suggestions provided by the anonymous reviewers, which helped improve the quality of the manuscript.

References

[1]. Ramya C, Shreedhara KS (2017) A new PSO methodology for web documents retrieval. In: Proceedings of the International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), pp 1–5. https://doi.org/10.1109/ICEECCOT.2017.8284622

[2]. Bhopale AP, Tiwari A (2020) Swarm optimized cluster based framework for information retrieval. Expert Syst Appl 154:113441. https://doi.org/10.1016/j.eswa.2020.113441

[3]. Ito A, Uno Y, Masumura R, Ito M, Makino S (2009) Relevant document retrieval using a spoken document. In: 9th International Symposium on Communications and Information Technology, pp 1483–1488. IEEE.

[4]. Molgaard LL, Jorgensen KW, Hansen LK (2007) CastSearch: Context based spoken document retrieval. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol 4, pp IV-93–IV-96. IEEE.

[5]. Gasser R, Rossetto L, Schuldt H (2019) Towards an all-purpose content-based multimedia information retrieval system. arXiv:1902.03878.

[6]. El-Gayar MM, Mekky NE, Atwan A, Soliman H (2019) Enhanced search engine using proposed framework and ranking algorithm based on semantic relations. IEEE Access 7:139337–139349. https://doi.org/10.1109/ACCESS.2019.2942872

[7]. Kim J (2019) A document ranking method with query-related web context. IEEE Access 7:150168–150174. https://doi.org/10.1109/ACCESS.2019.2947333

[8]. Jin L, Feng L, Liu G, Wang C (2017) Personal web revisitation by context and content keywords with relevance feedback. IEEE Trans Knowl Data Eng 29(7):1508–1521. https://doi.org/10.1109/TKDE.2017.2685593

[9]. Diligenti M, Gori M, Maggini M (2004) A unified probabilistic framework for Web page scoring systems. IEEE Trans Knowl Data Eng 16(1):4–16. https://doi.org/10.1109/TKDE.2004.1264818

[10]. Wang X, Macdonald C, Tonellotto N, Ounis I (2021) Pseudo-relevance feedback for multiple representation dense retrieval. In: Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval, pp 297–306.
[11]. Ren SH (2013) Relevant document retrieval via discrete stochastic optimization. In: International Conference on Computer, Wireless Communication and Measurement Technology, pp 73–76. https://doi.org/10.1109/ICCWAMTIP.2013.6716603
[12]. Fan W, Pathak P, Zhou M (2009) Genetic-based approaches in ranking function discovery and optimization in information retrieval—A framework. Decis Support Syst 47(4):398–407. https://doi.org/10.1016/j.dss.2009.04.005
[13]. Sharma DK, Pamula R, Chauhan DS (2024) A hybrid evolutionary algorithm based automatic query expansion for enhancing document retrieval system. J Ambient Intell Humaniz Comput 15(1):829–848. https://doi.org/10.1007/s12652-022-04395-2
[14]. Drias H (2011) Parallel swarm optimization for web information retrieval. In: Third World Congress on Nature and Biologically Inspired Computing, pp 249–254. https://doi.org/10.1109/NaBIC.2011.6089605
[15]. Djenouri Y, Belhadi A, Belkebir R (2018) Bees swarm optimization guided by data mining techniques for document information retrieval. Expert Syst Appl 94:126–136. https://doi.org/10.1016/j.eswa.2017.10.042
[16]. Ganesan S, Sivaneri A, Selvaraju S (2014) Evolving interest based user groups using PSO algorithm. In: International Conference on Recent Trends in Information Technology (ICRTIT), pp 1–6. https://doi.org/10.1109/ICRTIT.2014.6996196
[17]. Alam S, Dobbie G, Koh YS, Riddle P (2014) Web bots detection using particle swarm optimization based clustering. In: IEEE Congress on Evolutionary Computation (CEC), pp 2955–2962. https://doi.org/10.1109/CEC.2014.6900476
[18]. Kumar S, Kumar R (2021) WDPMA: An MA-based model for web documents prioritization. Int J Inf Technol Web Eng 16(2):1–24. https://doi.org/10.4018/IJITWE.2021040101
[19]. Wang Z, Sun X (2008) An efficient web query optimization algorithm based on LDA and MA. In: International Conference on Multimedia and Information Technology, pp 50–53. https://doi.org/10.1109/MMIT.2008.31
[20]. Zhao D, Liu L, Yu F, Heidari AA, Wang M, Chen H, Muhammad K (2022) Opposition-based ant colony optimization with all-dimension neighborhood search for engineering design. J Comput Des Eng 9(3):1007–1044. https://doi.org/10.1093/jcde/qwac034
[21]. Kumar S, Singh HK, Tripathi B (2023) A hybrid GA-FF based ranking method for reorganizing website structure. In: 6th International Conference on Information Systems and Computer Networks (ISCON), pp 1–6. https://doi.org/10.1109/ISCON57294.2023.10112045
[22]. Lin CC (2006) Optimal Web site reorganization considering information overload and search depth. Eur J Oper Res 173(3):839–848. https://doi.org/10.1016/j.ejor.2005.01.046
[23]. Samanta D, Dutta S, Galety MG, Pramanik S (2022) A novel approach for web mining taxonomy for high-performance computing. In: Cyber Intelligence and Information Retrieval: Proceedings of CIIR 2021, pp 425–432. Springer, Singapore. https://doi.org/10.1007/978-981-16-9447-9_36

[24]. Ehikioya SA, Zeng J (2021) Mining web content usage patterns of electronic commerce transactions for enhanced customer services. Eng Rep 3(11):e12411. https://doi.org/10.1002/eng2.12411
[25]. Vyas C (2019) Evaluating state tourism websites using Search Engine Optimization tools. Tour Manag 73:64–70. https://doi.org/10.1016/j.tourman.2018.12.019
[26]. Singh H, Kaur P (2021) ACO with heuristic desirability for Web page positioning problem. In: Malik H, Iqbal A, Joshi P, Agrawal S, Bakhsh FI (eds) Metaheuristic and Evolutionary Computation: Algorithms and Applications. Stud Comput Intell, vol 916. Springer, Singapore. https://doi.org/10.1007/978-981-15-6407-5_13
[27]. Abdulmunem ME, Naamha EQ (2023) Link analysis in Web information retrieval: A survey. Iraqi J Sci 64(9):4747–4766. https://doi.org/10.24996/ijs.2023.64.9.35
[28]. Panicker S, Vijay Kumar TV, Yadav D (2024) A novel transformation-based e-commerce website structure optimization model. IETE Tech Rev 41(1):3–12. https://doi.org/10.1080/02564602.2022.2150327
[29]. Bessghaier N, Soui M, Ghaibi N (2022) Towards the automatic restructuring of structural aesthetic design of Android user interfaces. Comput Stand Interfaces 81:103598. https://doi.org/10.1016/j.csi.2022.103598
[30]. Amati G, Van Rijsbergen CJ (2002) Probabilistic models of information retrieval based on measuring the divergence from randomness. ACM Trans Inf Syst 20(4):357–389. https://doi.org/10.1145/582415.582416
[31]. Jalilifard A, Caridá VF, Mansano AF, Cristo RS, da Fonseca FPC (2021) Semantic sensitive TF-IDF to determine word relevance in documents. In: Advances in Computing and Network Communications: Proceedings of CoCoNet 2020, Vol 2, pp 327–337. Springer, Singapore. https://doi.org/10.1007/978-981-16-6285-0_27
[32]. Oleinik A (2022) Relevance in Web search: Between content, authority and popularity. Qual Quant 56(1):173–194. https://doi.org/10.1007/s11135-021-01140-4
[33]. Xiong H, Bian J, Li Y, Li X, Du M, Wang S et al (2024) When search engine services meet large language models: Visions and challenges. IEEE Trans Serv Comput. https://doi.org/10.1109/TSC.2024.3366215
[34]. Gampa P, Fujita S (2021) BanditRank: Learning to rank using contextual bandits. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), pp 259–271. Springer, Cham. https://doi.org/10.1007/978-3-030-75768-7_21.