iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 5 **Issue:** XI **Month of publication:** November 2017

**DOI:** http://doi.org/10.22214/ijraset.2017.11117

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887*
*Volume 5 Issue X1, November 2017- Available at www.ijraset.com*

# Comparing Performance of Various Optimization Algorithms for Effective Information Retrieval – A Review

Tahera Shaikh[1], Shiwani Gupta[2]
[1, 2] *Computer Engineering Department, Mumbai University*

***Abstract*: *Information Retrieval (IR) is finding documents of an unstructured source like text that fulfills an information need from within large collections usually stored on computers. Because of the semantic disconnect between query and documents, IR is accountable to return a lot of unwanted items. There is a rapid growth of the amount of data available in electronic libraries, through Internet and enterprise network mediums, advanced methods of search and information retrieval are in demand. Information retrieval systems, designed for storing, maintaining and searching large-scale sets of unstructured documents, are the subject of intensive investigation.***

***The researchers have introduced various optimization algorithms in the field of Information Retrieval like Swarm Intelligence and Genetic Algorithm. These algorithms are inspired by nature to solve the IR problems. They have used these algorithms on different sets of databases. There is clear demand for fine tuning the performance of IR on Web data. The paper compares the various optimization algorithms like Accelerated Particle Swarm intelligence algorithm (APSO) and Bat algorithm, Ant Colony Optimization (ACO), Artificial Bee Colony (ABC) and Genetic Algorithm (GA).***

***Keywords: Information Retrieval; optimization; Accelerated particle swarm intelligence algorithm; Bat algorithm; Query Expansion***

## I. OBJECTIVE

The objective of this paper is to compare various optimization algorithms in information retrieval domain. The researchers have used various optimization algorithms to solve the problems for information retrieval. Here we have studied various information retrieval optimization algorithms. Different algorithms have been implemented on different sets of databases like MEDLINE database, Web database etc. We are comparing these algorithms based on effectiveness.

## II. INTRODUCTION

### *A. Information Retrieval*

During the last decade the information over the web have increased and optimization of information retrieval effectiveness has driven the quality of the results over the web, People are more trusting and preferring web search as a source of information. Information retrieval has come out of academic discipline to become the basis of most preferred and reliable source of information.
The field of information retrieval began with scientific library records and scientific publications; it spread rapidly in other domains like journalism, lawyers and medical fields. Information retrieval then spread in web information access. The information retrieval provides solution in finding relevant information in unstructured information [5].

### *B. Optimization (Swarm Intelligence)*

The researchers have used number of optimization technique in information retrieval domain. There are various models of IR and methods for optimization. Here we are concentrating on one of stochastic optimization technique called swarm intelligence. Swarm intelligence is the study of computational systems inspired by the 'collective intelligence'. Collective Intelligence emerges through the working together of large numbers of similar agents in the environment. Schools of fish, flocks of birds, and colonies of ants are some the examples. The property of swarm intelligence is self-organization, decentralization and distribution throughout the environment. The problems are solved in nature like foraging for food, prey evading, and colony relocation through SI. The information is stored and transferred by the means of agents such as proximity in fish and birds, pheromones in ants and dancing in bees [1] [6].

©IJRASET (UGC Approved Journal): All Rights are Reserved

## III. LITERATURE SURVEY

### A. Accelerated Particle Swarm Optimization (APSO) [1]

In this paper [1], the author has used swarm intelligence algorithm called Accelerated Particle Swarm Optimization (APSO) to solve problem of Query Expansion (QE). When the query is expanded, the numbers of expanded candidate keywords are generated. Here, to select the best and most relevant keyword is very complex. The author has used APSO to solve this problem. They have used three steps to achieve it which are:

*1)* Retrieve the pseudo-relevant documents to the initial query
*2)* Extract the best expansion keywords from the pseudo relevant documents to constitute the expanded query
*3)* Retrieve the relevant documents to the expanded query.

### B. Ant Colony Optimization (ACO) [3]

In this paper [3], the authors have used multiple techniques to employ Ant colony. They have used Naïve Rank, Random Rank, and Session Rank to implement different functionality of Ant Colony. Ranking results of diverse types of queries depends on users' intent. Queries have been clustered in three distinct categories on the basis of user's intent namely; Informational queries, Navigational queries, Transactional queries and they have used different algorithms to address these queries. The famous family of Ant Colony Optimization (ACO) algorithms is inspired by the Ants swarm intelligence and their use of stigmergic processes.

### C. Artificial Bee Colony (ABC) [5]

In this Paper [5], here the model shows collective intelligence of honeybee. It consists of three essential components: food sources, employed foragers, and unemployed foragers. They have used Page Rank algorithm to implement this technique. They have implemented it in three steps as follows: (i)

*1)* Calculation of user interest
*2)* Growth analysis Rate
*3)* Total site linking

### D. Bat algorithm (BA) [6]

In this paper [6], they have used swarm intelligence to solve the problem of query expansion, First, they have calculated rank of the pseudo relevant documents using document scoring function like OKAPI 25, next they have extracted keyword candidate using term scoring function e.g. RSJ, Rocchio, next they have put the best ranked keyword in original query, and finally the document is retrieved by using document scoring function.

### E. Genetic Algorithm (GA) [2]

In this paper [2], they have used genetic algorithm which creates a set of Apache Lucene search queries for text document clustering. The document which is added in only one cluster will add strength of the cluster, whereas the document which is added in more than one cluster will decrease the suitability. Here extra labelling step is not necessary because the final search queries are effortlessly understood and present cluster in uncomplicated manner

## IV. COMPARISON RESULTS

The following table states the merits and demerits of various optimization techniques based on swarm intelligence algorithm. The table shows the comparative analysis of the existing methodologies; it can help for the researcher for further research.

TABLE I.
MERITS AND DEMERITS OF INFORMATION RETRIEVAL OPTIMIZATION ALGORITHMS

| Title | Merits | Demerits |
|---|---|---|
| An accelerated PSO for query expansion in web information retrieval: application to medical dataset [1] | Claims to be better than other swarm intelligence algorithms like bat algorithm and firefly algorithm. It is a combinatorial optimization | Only implemented on MEDLINE<br>Authenticity should be checked with other types of database |
| Document clustering with evolved search queries [2] | The final search queries are easily understood and offer a simple explanation of the clusters, meaning that an extra cluster labeling stage is not required | It is often the case that the ideal number of clusters is not known in advance<br>Only dealt with the simple case where a document can only belong to one cluster |

©IJRASET (UGC Approved Journal): All Rights are Reserved

| | | |
|---|---|---|
| An ant-colony based approach for real-time implicit collaborative information seeking [3] | Ranking results of diverse types of queries depending on users' intent. | Only tested with three different algorithms to make ant colony approach |
| Implementation of improved IR System through Swarm Intelligence Technique [4] | They have used most of the swarm algorithm for information retrieval<br>It is implementable on large datasets | All the search queries generated here use a simple OR between terms. |
| Artificial Bee Colony approach for ranking web pages [5] | Claims to be more efficient compared with tradition page Rank Algorithm | Authenticity should be checked with other types of real time database |
| Bat algorithm for efficient query expansion: Application to MEDLINE [6] | It is a combinatorial optimization<br>The bat algorithm aims to extract the best expansion keywords, in one hand and to select the best relevant documents to the expanded query on other hand | The performance of RSJ and Rocchio is better than the proposed approach<br>Only implemented on MEDLINE |

The following table states the brief summary of the research work. The table also shows the future scope which authors have stated and the gap analysis for the further studies. These will help in learning the topic thoroughly.

Table II.
RESEARCH TRENDS IN INFORMATION RETRIEVAL OPTIMIZATION

| Title | Summary | Gap Analysis | Future Scope |
|---|---|---|---|
| An accelerated PSO for query expansion in web information retrieval: application to medical dataset [1] | They used application of the APSO technique to efficiently solve the problem of Query Expansion (QE) in Web Information Retrieval (IR). | Different ranking and indexing approaches can be researched to improve the optimization.<br>They have used only one kind of dataset the approach can be broadened by using same algorithm on web database and twitter. | To test and experiment the proposed approach on other existing Web datasets.<br>To design and implement a parallel version of the APSO in order to further improve both retrieval effectiveness and computational efficiency. |
| Document clustering with evolved search queries [2] | They defined a system whereby an island model genetic algorithm (GA) generates entities which can generate a set of Apache Lucene search queries for text document clustering. | Different query formats can be tested to increase precision without altering the readability of the queries. | In text clustering problems, it is often the case that the perfect number of clusters is not known in advance. The future scope is to experiment with introducing an extra gene in the chromosome used to specify the number of clusters |
| An ant-colony based approach for real-time implicit collaborative information seeking [3] | They designed and developed three different algorithms that employ an ACO-inspired strategy to provide implicit collaborative- seeking features in real time to search engines | Different ranking and indexing approaches can be researched to improve the optimization.<br>Can be researched with different sets of algorithms for optimization. | To blend some concepts explored by existing ACO extensions in our model, such as limiting the amount of deposited pheromone to avoid deposited as in Max-Min Ant System. |

©IJRASET (UGC Approved Journal): All Rights are Reserved

| | | | |
|---|---|---|---|
| Implementation of improved IR System through Swarm Intelligence Technique [4] | They have used different optimization techniques or algorithms, and with the help of those algorithms the datasets are first reduced and then that datasets are provided as an input to the algorithms i.e. Particle Swarm Optimization, Ant Colony Optimization | Different High dimensional datasets can be tested to prove the results. | Optimization of information retrieval system, hence forming an improved IR system on high dimensional textual datasets. |
| Artificial Bee Colony approach for ranking web pages [5] | In This Paper Artificial Bee Colony (ABC) approach is used as a new method for web mining particularly in ranking web pages. | The information gathered for analysis can be increased for more accurate results. | The algorithm can be adopted by any Search Engine and to implement it, it can be extended for different datasets. |
| Bat algorithm for efficient query expansion: Application to MEDLINE [6] | Bat-Inspired Approach is used to improve the computational cost. Then, this approach is used to retrieve both the best expansion keywords and the best relevant documents simultaneously. | Different ranking and indexing approaches can be researched to improve the optimization. | To test and experiment the proposed approach on other existing Web datasets. |

## V. CONCLUSIONS

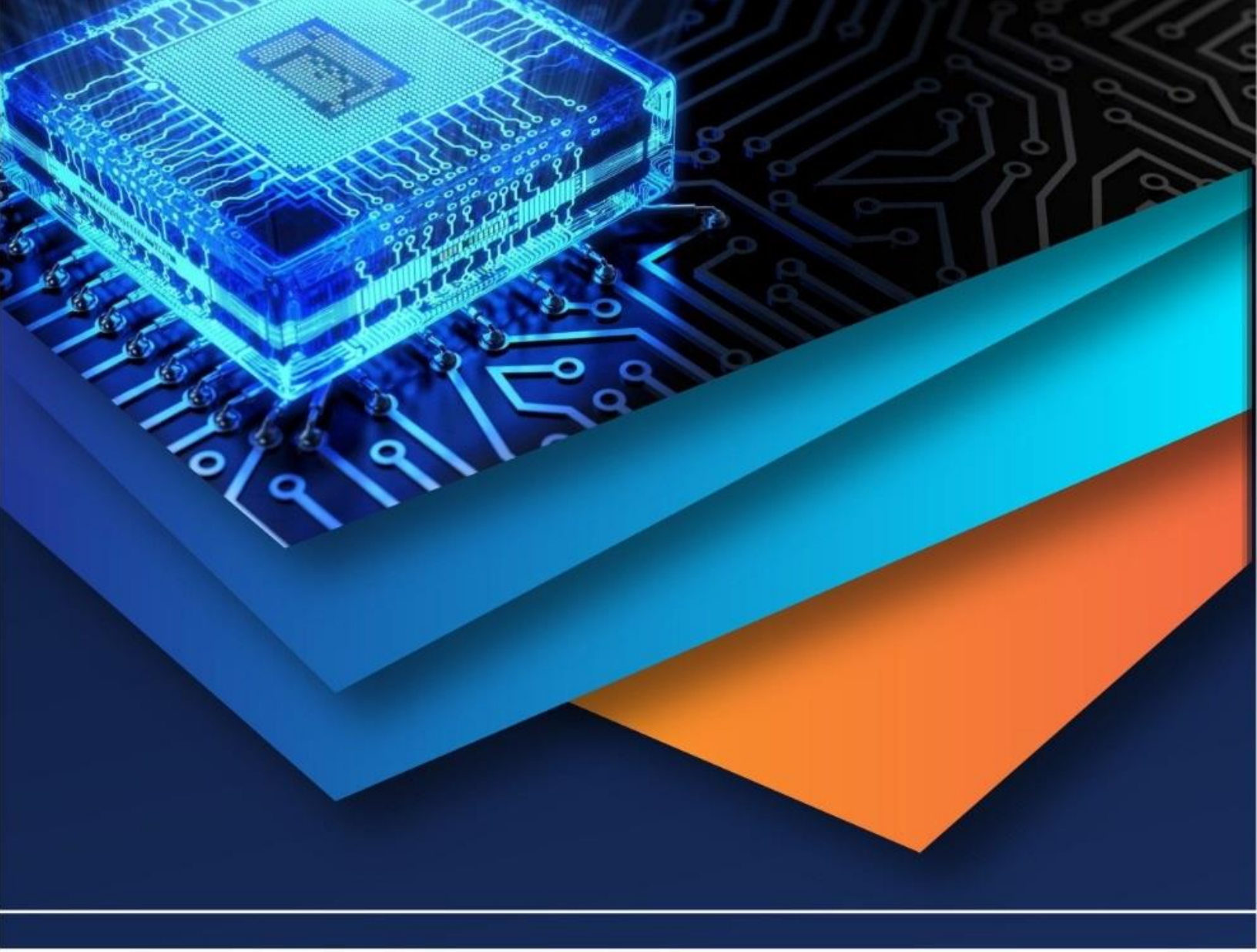
The authors have used various techniques to solve the problem of information retrieval, they have used optimization algorithms like swarm intelligence algorithms, there are various kinds of swarm intelligence algorithms like accelerated particle swarm intelligence algorithm (APSO) and Bat algorithm, Ant Colony Optimization (ACO), Artificial Bee Colony (ABC), Genetic Algorithm.

This work compared various researches on optimization algorithms, with their future scope and gap analysis. This paper also shows comparative study of these optimization algorithms. The author described the description and methodology used for optimization algorithms and the paper also mentions demerits to be considered before doing any implementation and research on mentioned algorithms.

## REFERENCES

[1] Ilyes Khennak, Habiba Drias, "An accelerated PSO for query expansion in web information retrieval: application to medical dataset", Springer Science + Business Media, pp 1-16, International journal of Applied Intelligence, 2017.
[2] L.Hirsch, A. D. Nuovo, "Document clustering with evolved search queries" IEEE congress on Evolutionary Computation 2017, Spain.
[3] A. Malizia, K. A. Olsen, T. Turchi, P. Crescenzi, "An ant-colony based approach for real-time implicit collaborative information seeking", Information Processing and Management. Vol 53(3) May 2017. 608-623, ScienceDirect, Elsevier. DOI: 10.1016/
[4] S. Lohar, P. Howale, S. Pradhan, M. Redekar, "Implementation of improved IR System through Swarm Intelligence Technique", IJRITCC, Oct 2015.
[5] G. Anuradha, G. Lavanya, "Artificial Bee Colony approach for ranking web pages", IJCA, Aug 2014.
[6] Khennak I, Drias H, "Bat algorithm for efficient query expansion: Application to MEDLINE", Proceedings of the 4th World Conference on Information

©IJRASET (UGC Approved Journal): All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)