

# A Comparison and Analysis of various extended Techniques of Query Optimization

<sup>1</sup>Atul Garg, <sup>2</sup>Dimple Juneja

<sup>1</sup>M. M. Institute of Computer Technology & Business Management (MCA)  
Maharishi Markandeshwar University, Mullana-Ambala, Haryana, 133201, India

<sup>2</sup>Technology Education and Research Integrated Institutions  
Kurukshetra University, Kurukshetra, Haryana, India

## Abstract

This paper presents the detailed comparison of various evolutionary algorithms developed for query optimization on web servers. It is very much a known fact now that information on the web has been growing exponentially and the need of efficient extraction of information was felt long back. Therefore, researchers have been putting their time and efforts for developing various algorithms that suit to the need of changing times. Development of these evolutionary algorithms has been motivated from biological and social behaviour of animals, birds and human beings. The work aims to compare the existing works with the intention to find to scope of improvement amongst these, if any.

**Keywords:** *Query Optimization, Genetic Algorithms, Memetic Algorithm, Particle Swarm Optimization, Ant Colony Optimization, Shuffled Frog Leap Algorithm.*

## 1. Introduction:

The Web is continuously attracting waves of new users and service providers. It is now the de facto medium for exchanging information and services. This globalization has also spurred the development of tools and aids to navigate and share information in corporate intranets that were previously accessible online only in a restrictive way and at prohibitive costs. The information age revolution has highlighted the role of the database management system (DBMS) as a key enabling technology. DBMSs are currently the technology of choice for modelling, storing, managing, and efficiently querying large amounts of information. They also provide functions for protecting data integrity, facilitating reliable and flexible data access and manipulation, synchronizing concurrent accesses from applications, and securing data[7]. Accordingly, in order to make use of this vast amount of data, efficient techniques to retrieve web document information based on its content were needed to be developed. As a result, the role of information retrieval (IR) systems became more important. One of the most important and difficult operations in information retrieval was to generate queries that can succinctly identify relevant documents and reject irrelevant documents [6].

Further, one of the major aspects that have caught the attention of the emerging researchers was the way in which the users interact with the structure and content of web. Forth at purpose, multiple analyses and models were generated to understand web user behaviour in order to display relevant content and maximize traffic [8]. In continuation, the researchers have proposed evolutionary-based algorithms for searching near-optimum solutions to mathematical problems. Evolutionary algorithms (EAs) [19] are stochastic search methods that mimic the metaphor of natural biological evolution and/or the social behaviour of species. For examples, how ants find the shortest route to their destination (food) and how birds find their destination during migration. Many computational systems such as genetic algorithms (GAs), memetic algorithms (MAs), particle swarm optimization (PSO), ant colony systems (ACO) and shuffled

frog leaping (SFL) seeking fast and robust solutions to complex optimization problems have been developed so far.

In this paper, a comparative analysis of above mentioned algorithms is being presented with the intention to explore future scope of research and improvement. The parameters and performance comparison among these algorithms is presented in tabular form. The remainder of the paper is organized as follows: Section 2 introduces the related work. Section 3 presents the analyses in detail and compare these algorithms. The paper ends with conclusion and references.

## 2. Related Work

With the exponential expansion of the data on web servers every day, it was observed that the type of data is also varying from static to dynamic data. The data formats are now being changed to make web more secure. Thus, the need for tools to maximize the portability, reusability and interoperability of arbitrary computing services was felt leading to the many evolutionary works.

Hawash et. al [1] proposed two disk based versions of trace equivalence and bi-similarity algorithms for summarizing data graphs. Authors adapted the algorithms to the RDF data model. They used memory-based version of the algorithm to summarize directed labelled graphs but proved to be inefficient if applied to large data graphs because of memory limitations. Their work introduced a scalable disk-based version of the algorithm to reduce large graphs. The analysis of conducted experiments on both algorithms using relatively large RDF datasets proved that the two algorithms are scalable and totally memory independent.

Srivastava et. al [2] have proposed the overall goal of a general-purpose web service management system (WSMS), enabling clients to query a collection of web services in a transparent and integrated fashion. The focus was on new query optimization issues that arose in a WSMS. Their execution model consists of pipelined query processing over web services, and they derive the “bottleneck” cost metric to characterize the cost of a pipelined plan. For this cost metric, they have devised new algorithms not only to decide the optimal arrangements of web services in a pipelined plan, respecting precedence constraints but also to decide the optimal chunk size to use when sending data to each web service. But the algorithms in this paper formed the basis of a WSMS query optimizer.

Beran et. al [3] presented a multi-layered blackboard approach for database query optimization in heterogeneous environments that eases the way of handling not only query related optimization problems but also covers data and resource-specific aspects as well, by assigning costs and rating decisions made in the query optimization process. According to them, distributed queries heavily depend upon the quality of the constructed optimal query execution plan. Their evaluation model could prove even small modifications in the structure of a QEP leading to benefits in the query optimization.

Tang et. al [4] proposed a constraint-based optimization framework. Specifically, the expertise matching problem was transformed to a convex cost flow problem and the objective was then to find a feasible flow with minimum cost under certain constraints. According to the authors, the framework could achieve an optimal solution. They conducted experiments on two different genres of tasks namely, conference paper-reviewer assignment and course-teacher assignment. Experimental results validated the effectiveness and efficiency of their proposed approach.

According to Sharma et. al [5] one major problem of irrelevant data on Internet is lack of user knowledge in framing. They proposed a query log analysis to implement effective web search. The result optimization method was based on user's feedback, which determined the relevance between web pages and user query words. The returned pages with improved page ranks were directly mapped to the user feedbacks and dictated higher relevance than pages that exist in the result list but are never accessed by the user. The results obtained from practical evaluation are quite promising in respect to reduced search space and improved effectiveness of interactive web search engines.

Wang and Sun [6] proposed an efficient document query algorithm based on LDA and memetic algorithm (MA). The authors investigated the capability of the memetic algorithm (MA) boosted by LDA-based dimensionality reduction method for Web document query optimization in the context of information retrieval. A series of experiments were conducted to compare the MA-based algorithm with other document query optimization algorithm based on relevant feedback (RF) approach and genetic algorithm (GA). According to their comparison experiments MA is more effective than GA and RF because the LDA-based dimensionality reduction method reduces the cost of the post-processing involved in document retrieval, and improves retrieval speed and efficiency.

An engraved look at the above literature highlights the fact that researchers have been putting efforts towards developing algorithms for web query optimization. The existing algorithms have one or the other advantage over each other. Next section discusses the working, pros and cons of five important algorithms mentioned above.

### **3. Analysis and Comparison**

Information retrieval (IR) is a used for storage, maintenance and search large amounts of data available on various servers using Internet. Now a day, the data on web could be textual, visual, audio or multimedia documents etc. The amount of documents contained in data collections managed by IRS (Information Retrieval System) is usually very large and the task of easy, efficient and accurate information search is specially challenging. The researchers have proposed various other techniques and algorithms for searching near-optimum solutions inspired by different natural processes such as genetic algorithms (GAs), memetic algorithms (MAs), particle swarm optimization (PSO), ant colony systems (ACO) and shuffled frog leaping (SFL), just to name a few.

#### **3.1 Genetic Algorithm**

GAs is inspired by biological system's improved fitness through evolution [11]. Genetic algorithms are based on selection, crossover and mutation. It evolves a population of chromosomes representing potential problem solutions encoded into suitable data structures [14]. It has the potential for solving intractable optimization problems [12] and work with random population of solutions (chromosomes). Genes holds a set of values for the optimization variables [11, 12, 13, 14, 16]. To simulate the natural survival of the fittest process, best chromosomes exchange information (through crossover or mutation) to produce offspring chromosomes. The offspring solutions are then evaluated and used to evolve the population if they provide better solutions than weak population members as shown in figure 1. Usually, the process is continued for a large number of generations to obtain a best-fit (near optimum) solution. GAs is no doubt a very promising technique for solving optimization problems with a trade-off between speed and perfection. But, it can't be implemented on all

problems such as electric circuit problem, travelling salesman problem, multiple object problems etc.

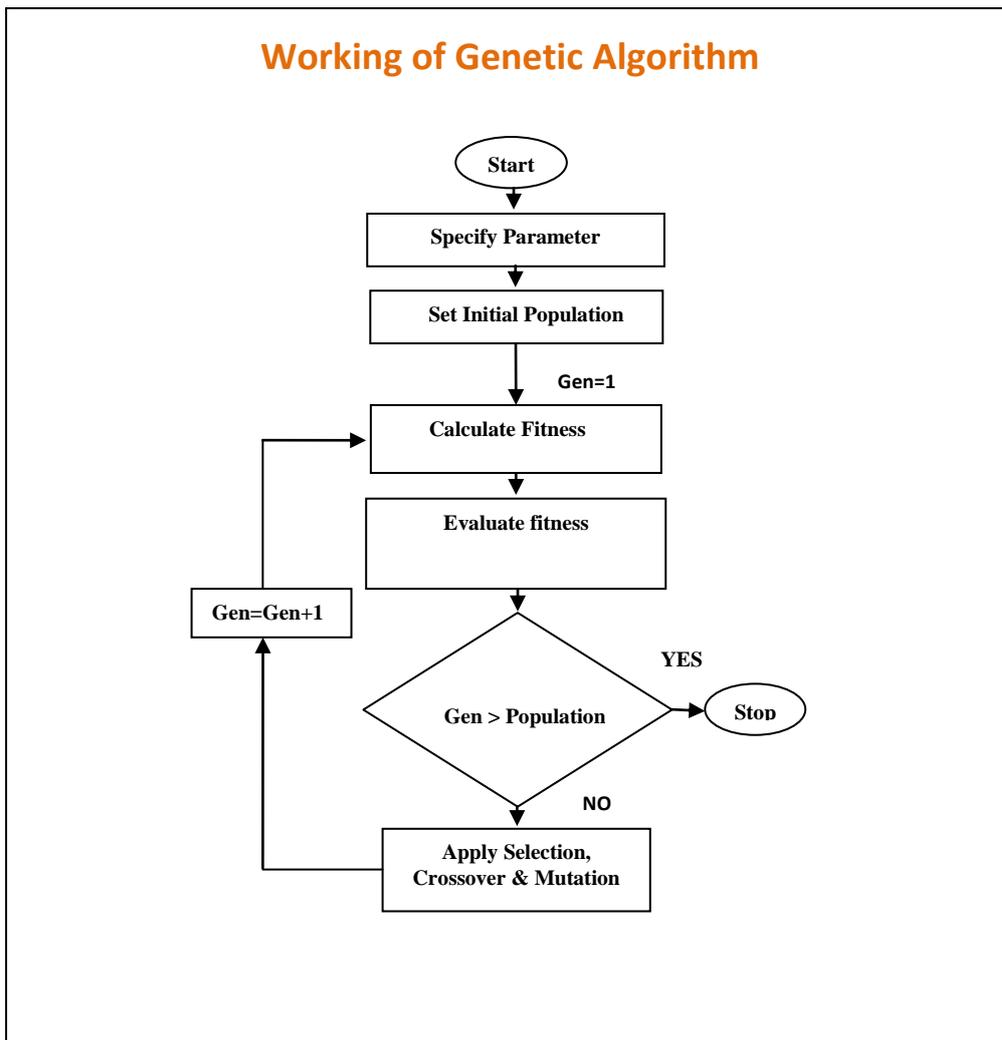


figure 1: Working of Genetic Algorithms

### 3.2 Memetic Algorithms

MAs are inspired by the notion of evolution and inspired by Richard Dawkin's notion of a meme [20, 21]. MAs are similar to GAs but the elements that form a chromosome are called memes, not genes. Meme was defined as a noun that communicates the idea of a unit of cultural transmission. MAs are different from GAs as they incorporate local search process to refine the solutions before they get involved in the evolutionary process. Memes are analogous to genes as they are self replicating, but they differ from genes as they are transmitted through imitation rather than being inherited. The unique aspect of the MAs algorithm is that all chromosomes and offsprings are allowed to gain some experience, through a local search [figure 2], before being involved in the evolutionary process [6, 13]. A quotation of Dawkins [20] about MAs is:

“Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation.”

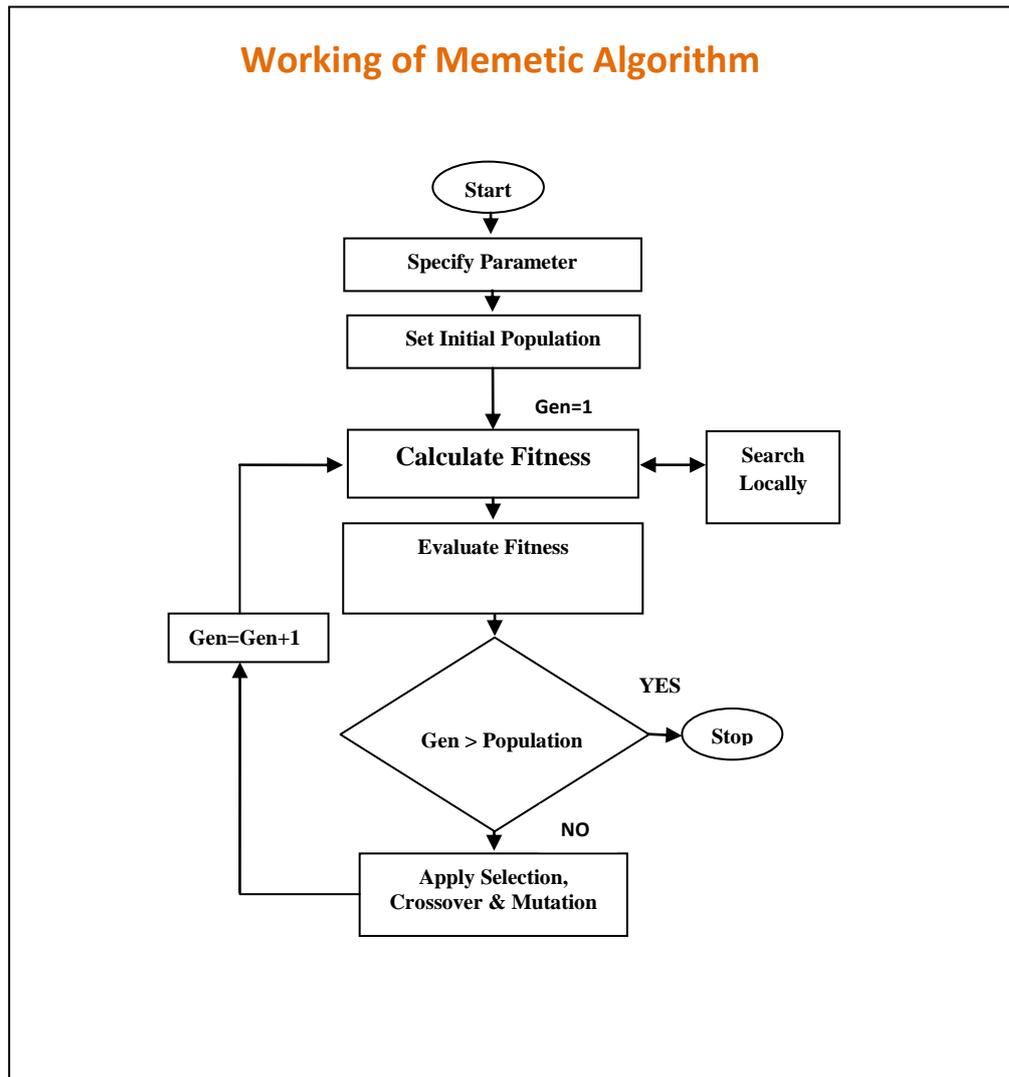


figure 2: Working of Memetic Algorithms

### 3.3 Particle Swarm Optimization

PSO was developed by Kennedy and Eberhart [15]. PSO is based on the behaviour of a flock of migrating birds trying to reach an unknown destination. In PSO, each solution is a 'bird' in the flock and is referred to as a 'particle'. A particle is analogous to a chromosome (population member) in GAs. Physically, birds look in a specific direction (towards their destination) and during their communication, they identify the bird that is in the best location. Accordingly, each bird speeds towards the best bird using a velocity that depends on its current position. PSO algorithm had basic three steps, namely, generating particles' positions and velocities, velocity update, and finally, position update. Each bird investigates the search space from its new local position, and the process repeats until the flock reaches a desired destination [13, 16, 17] as shown in figure 3.

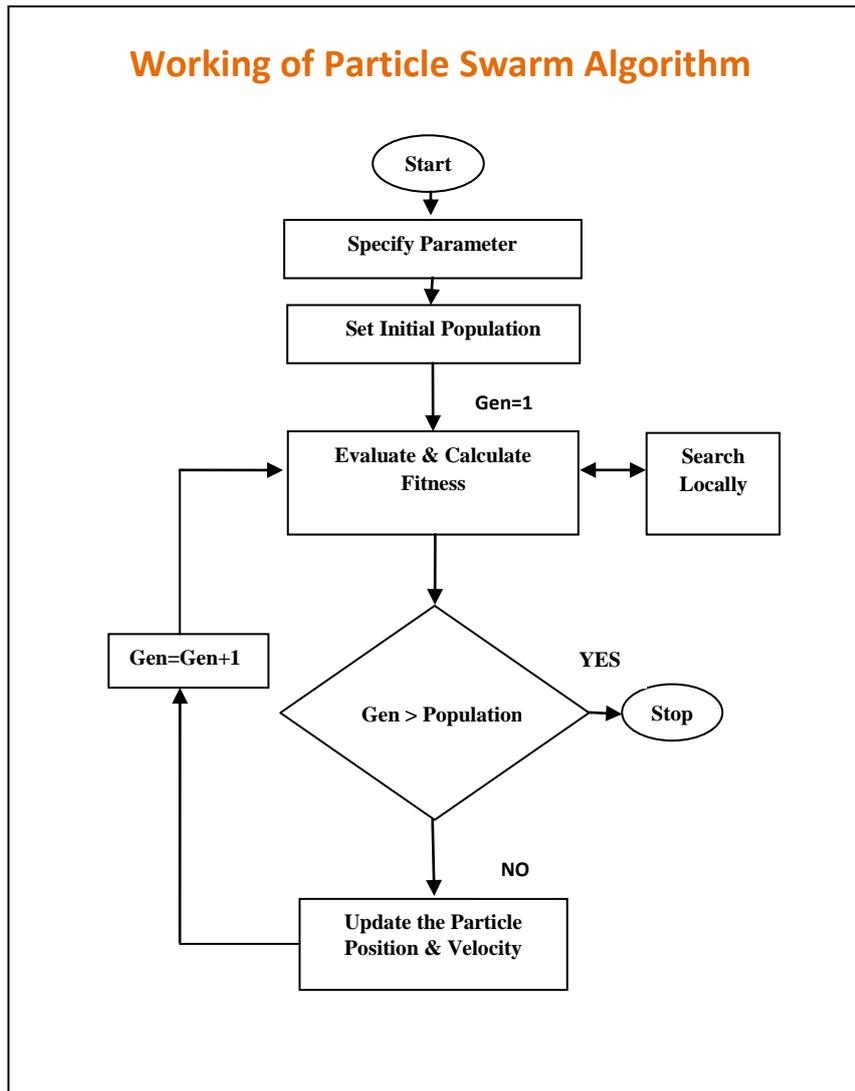


figure 3: Working of Particle Swarm Algorithms

### 3.4 Ant-Colony Optimization (ACO)

Ant Colony System (ACS) is an agent-based system, which is based on the biological ants and their social behaviour see Figure 4A. The Ants can choose any path to reach its destination. Initially it could be nearest or it could be longest. ACS was proposed by Dorigo et al. in 1997 for combinatorial optimization problems. This new approach is called Ant Colony Optimization (ACO). The working of ACO is being given in figure 4.

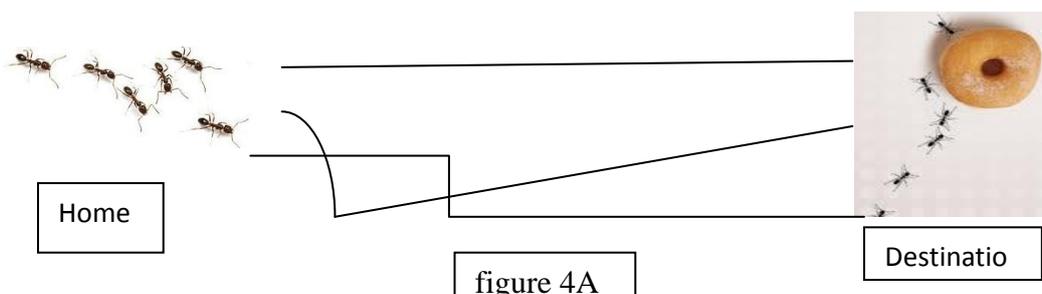


figure 4A

The Ants are best for travel sales man problem. And these inherit parallelism and works in a team. The ACO is best for dynamic application. The idea behind ACO is to produce a model to search for a minimum cost path. The behaviour of artificial ants is inspired from real ants. Each ant builds a path from its source node to its destination. Ants are able to find the shortest path for its source of food. These deposit pheromone trails in their travelling, as a communication with others. After carrying the food and start returning back, following their pheromone trails, and still depositing more pheromone. While an ant builds a path, it gets quantitative information about the path cost and qualitative information about the amount of traffic in the network. Then, another ant travelling through the same path will carry this information [13, 18]. The popular application of ACO is dynamic shortest path problems arising in telecommunication networks problems.

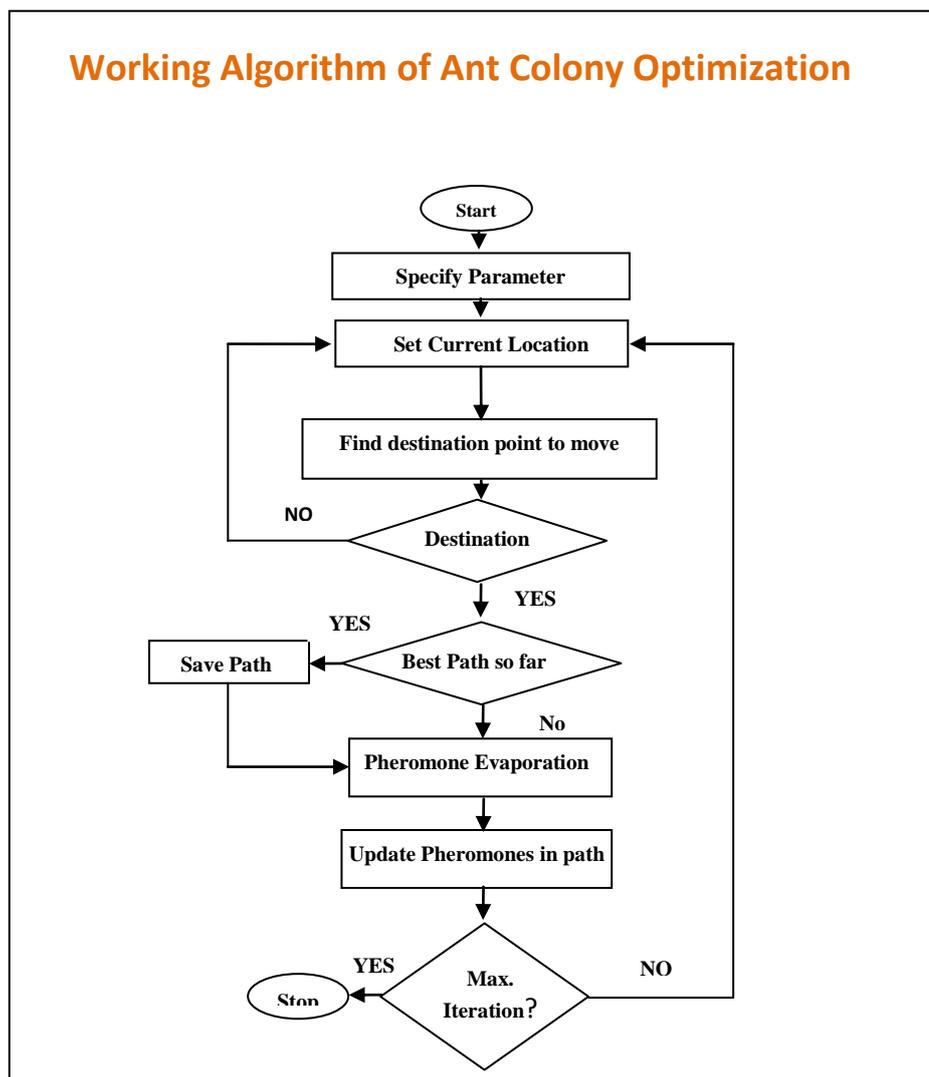
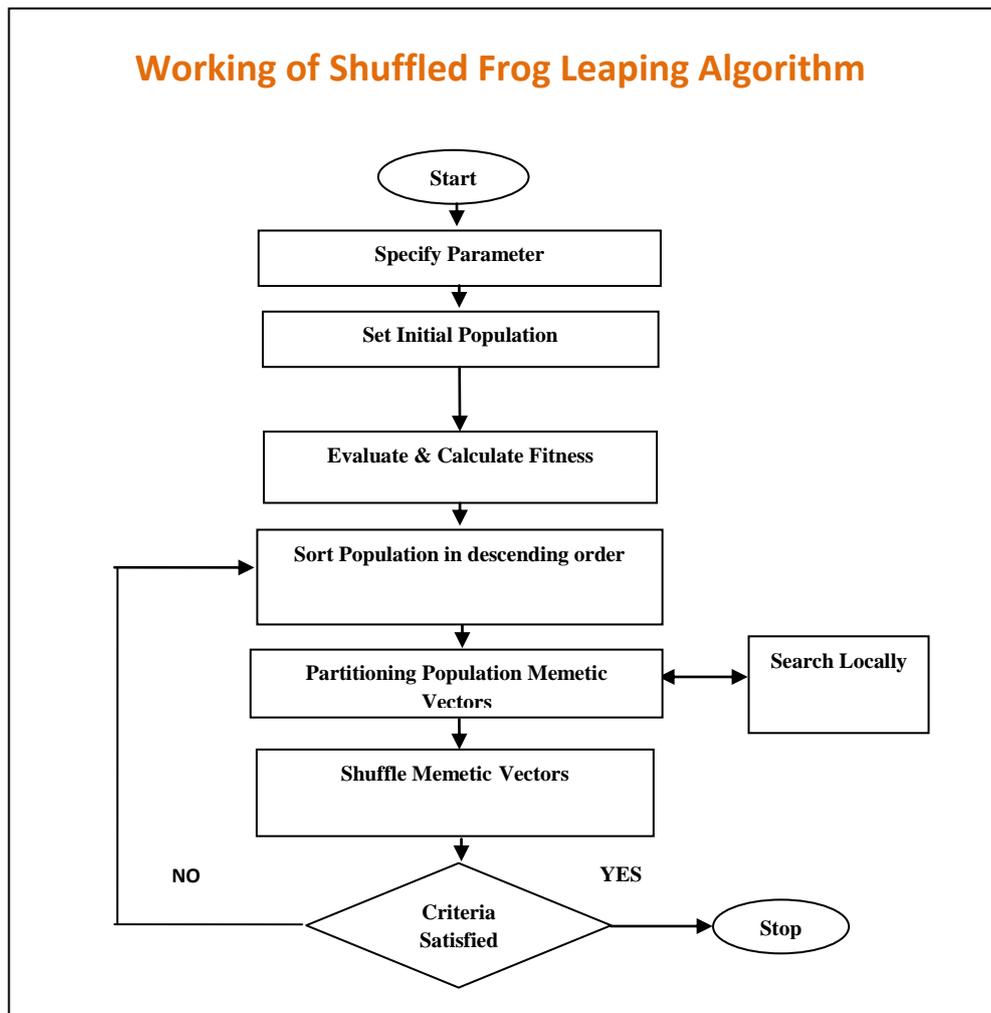


Figure 4: Working of Ant Colony Optimization

### 3.5 Shuffled Frog Leaping Algorithm (SFL)

The SFL algorithm selects the benefits of the MAs and the social behaviour-based PSO algorithms. Individual frogs are not so important; rather they are seen as hosts for memes and described as a memetic vector. Each meme consists of a number of memotypes. The memotypes represent an idea in a manner similar to a gene representing a trait in a

chromosome in a genetic algorithm [10]. The different memetic vector is considered as different cultures of frogs, each performing a local search as shown in figure 5. Within each memplex, the individual frogs hold ideas, that can be influenced by the ideas of other frogs, and evolve through a process of memetic evolution. After a defined number of memetic evolution steps, ideas are passed among memetic vector in a shuffling process [11, 13]. The local search and the shuffling processes continue until defined convergence criteria are satisfied [10].



**Figure 5: Shuffled Frog Leaping Algorithm**

As per above discussion, literature [6,9,10,11,12,13,14,16] depicts that each evolutionary algorithm has some common and some novel parameters. These parameters affect the processing time, quality and overall performance. Behaviour of GA and MA is based on genetic system, where PSO and ACO based on social behaviour than SFL is on both biological genetic and social behaviour. We analyzed that each algorithm is based on one or the other competitors e.g., MA is based on GA, PSO is based on MA and GA, ACO is on PSO and SFL is on MA and PSO. The performance of one is increased when the features of another algorithm are incorporated since the parameters and behaviour of each algorithm is different. These approaches provide good solutions regardless of the number of web objects to be arranged. A comparison table 1 highlighting the performance of each of the algorithms is shown below.

**Table 1: Comparative Analyses of Various Evolutionary Algorithms**

Sr. No.	Algorithm Manner	GA	MA		PSO	ACO	SFL
1.	<b>Behaviour</b>	Biological genetic system's	Biological genetic system's		Social behaviour	Social behaviour	Biological genetic & Social Behaviour
2.	<b>Parameters</b>	1.Population size 2.No of generations 3.Crossover rate 4. Mutation rate.	1. Population size 2.No of generations 3. Crossover rate 4. Mutation rate.	With Local Search Mechanism	1. Population size 2.Number of generation Cycles 3. The max. change of a particle velocity 4. Current Position	1.Number of ants 2.Number of iterations 3.Pheromone evaporation rate 4.Pheromone reward factor.	1. Number of frogs 2. Number of Memetic Vector 3. Number of generation for Each Memetic vector before shuffling 4.Number of shuffling iterations 5. Maximum step size.
3	<b>Replication</b>	Restricted	Less Restricted than GA		Not Restricted	Not Restricted	Not Restricted
4	<b>Transmission</b>	Parent to offspring	Any two individuals		Anyone in the group	-	-
5	<b>Locality</b>	Population based Search	Population based with Local Search		Local Search with position and velocity	Last Pheromone	Local Search
6	<b>Flexibility</b>	Not Flexible	Flexible		More Flexible than GA & MA	Flexible	More flexible than MA and PSO
7	<b>Method</b>	1. selection, 2. crossover 3. mutation	1.selection, 2.crossover 3.mutation	With Local search	1.Positions & velocities 2.velocity update 3.position update	Pheromone trails	Idea based shuffling process

As can be seen that although genetic and memetic algorithms are subjected to the same basic principles i.e. selection, crossover and mutation. Memetic algorithm is a much more flexible than genetic. Basic goals of genes and memes are different because they use different mechanisms to distribute information from one member of the population to another member. Genes can only be transmitted from parents to offspring whereas memes can be transmitted between any two individuals. Unlike genetic, if an improved idea is found in MA, PSO and SFL it can be adopted immediately than waiting for a full generation of genes. PSO does not incorporate survival of the best value, which features the removal of some candidate population members, individuals with lower fitness are removed with higher probability.

PSO is based on the intelligence. Like GA & MA, PSO don't have overlapping and mutation. The search can be carried out by the speed of the particle. The calculation in PSO is very simple than GA, MA and SFL. As Compare with GA, MA and SFL, it occupies the bigger optimization ability. PSO algorithm is totally based on coordination. ACO is efficient for Travelling Salesman Problem & can be used better in dynamic applications than GA and MA. Theoretical analysis is difficult in ACO. In ACO research is experimental rather than theoretical.

#### 4. Conclusions

In this paper, evolutionary-based search methods were presented. These include: GA, MA, PSO, ACO, and SFL. A brief description of each method is presented along with flowcharts to explain their flow and comparison table was presented in Table 1. For each algorithm initial setting of the parameters was established using values. As per analyzed from the literature PSO method was generally found to perform better than other algorithms in terms of success rate and solution quality, while being second best in terms of processing time.

#### References

- [1] Anton Deik, Bilal Faraj, Ala Hawash, Mustafa Jarrar "Towards Query Optimization for the Data Web - Two Disk-Based algorithms: Trace Equivalence and Bisimilarity". In proceedings of the International Conference on Intelligent Semantic Web – Applications and Services. Pages 131-137. ACM. ISBN 9781450304757. June 2010.
- [2] Utkarsh Srivastava, Kamesh Munagala, Jennifer Widom, Rajeev Motwani, "Query Optimization over Web Services", VLDB '06, September 1215, 2006, Seoul, Korea. Copyright 2006 VLDB Endowment, ACM 1595933859/06/09.
- [3] Peter Paul Beran, Werner Mach, Ralph Vigne, J'urgen Mangler and Erich Schikuta, "A Heuristic Query Optimization Approach for Heterogeneous Environments", 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, 978-0-7695-4039-9/10 \$26.00 © 2010 IEEE, DOI 10.1109/CCGRID.2010.65.
- [4] Wenbin Tang, Jie Tang, and Chenhao Tan, "Expertise Matching via Constraint-based Optimization", 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 978-0-7695-4191-4/10 \$26.00 © 2010 IEEE.
- [5] A.K. Sharma, Neha Aggarwal, Neelam Duhan and Ranjna Gupta, "Web Search Result Optimization by Mining the Search Engine Query Logs", 2010 International Conference on Methods and Models in Computer Science (ICM2CS-2010), 978-1-4244-9703-4/101\$26.00 ©2010 IEEE.
- [6] Ziqiang Wang, Xia Sun, "An Efficient Web Query Optimization Algorithm Based on LDA and MA", 2008 International Conference on MultiMedia and Information Technology, 978-0-7695-3556-2/08 \$25.00 © 2008 IEEE.

- [7] MOURAD OUZZANI , ATHMAN BOUGUETTAYA, “Query Processing and Optimization on the Web”, Distributed and Parallel Databases, 15, 187–218, 2004, \_c 2004 Kluwer Academic Publishers. Manufactured in The Netherlands.
- [8] Loyola, P., et al., Predicting web user behavior using learning-based ant colony optimization. Eng. Appl. Artif. Intel. (2011), doi:10.1016/j.engappai.2011.10.008.
- [9] Joglekar A, Tungare M. Genetic algorithms and their use in the design of evolvable hardware. <http://www.manastungare.com/articles/genetic/genetic-algorithms.pdf>; 2003, accessed on May 20,2004, 15 p.
- [10] Eusuff MM, Lansey KE. Optimization of water distribution network design using the shuffled frog leaping algorithm. J Water Resour Plan Manage 2003;129(3):210–25.
- [11] EMAD ELBELTAGI, TAREK HEGAZY and DONALD GRIERSON, “A modified shuffled frog-leaping optimization algorithm:applications to project management” , Structure and Infrastructure Engineering, Vol. 3, No. 1, March 2007, 53 – 60, ISSN 1573-2479 print/ISSN 1744-8980 online <sup>a</sup> 2007 Taylor & Francis.
- [12] A. Asllani, A. Lari, “Using genetic algorithm for dynamic and multiple criteria web-site optimizations”, European Journal of Operational Research 176 (2007) 1767–1777, 2005 Elsevier B.V. All rights reserved. doi:10.1016/j.ejor.2004.03.049.
- [13] EMAD ELBELTAGI, TAREK HEGAZY and DONALD GRIERSON, “Comparison among five evolutionary-based optimization algorithms” , Advanced Engineering Informatics 19 (2005) 43–53, q 2005 Elsevier Ltd. All rights reserved. doi:10.1016/j.aei.2005.01.004.
- [14] Pavel Kromer, Vaclav Snasel, Jan Platos and Ajith Abraham, “Evolutionary Improvement of Search Queries and Its Parameters”, 10th International Conference on Hybrid Intelligent Systems, ©2010 IEEE.
- [15] Kennedy J, Eberhart R. Particle swarm optimization. Proceedings of the IEEE international conference on neural networks (Perth, Australia), 1942–1948. Piscataway, NJ: IEEE Service Center; 1995.
- [16] Russell C. Eberhart and Yuhui Shi, “Comparison between Genetic Algorithms and Particle Swarm Optimization”, Evolutionary Programming VII, 1998 – Springer.
- [17] Rania Hassan Babak Cohanim† Olivier de Weck, “A COPMARISON OF PARTICLE SWARM OPTIMIZATION AND THE GENETIC ALGORITHM”, American Institute of Aeronautics and Astronautics  
[http://web.mit.edu/deweck/www/PDF\\_archive/3%20Refereed%20Conference/3\\_50\\_AI\\_AA-2005-1897.pdf](http://web.mit.edu/deweck/www/PDF_archive/3%20Refereed%20Conference/3_50_AI_AA-2005-1897.pdf)
- [18] Benjamín Barán and Rubén Sosa, “AntNet Routing Algorithm for Data Networks based on Mobile Agents”, Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. No. 12 (2001), pp 75-84.
- [19] Hartmut Pohlheim, “Evolutionary Algorithms: Overview, Methods and Operators”, GEATbx version 3.8, [www.geatbx.com](http://www.geatbx.com), December 2006.
- [20] Dawkins, R. (1976), The Selfish Gene. Oxford University Press.
- [21] Hema Banati1 and Shikha Mehta2, “A Multi-Perspective Evaluation of MA and GA for Collaborative Filtering Recommender System”, International journal of computer science & information Technology (IJCSIT) Vol.2, No.5, October 2010.