International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 15
May 2018

A Web-Based Crawler Using Bee Swarm Intelligent Algorithm

Ojoma S[1] , Onuodu F. E[2] , Nlerum P. A[3]
[1,2] Department Of Computer Science University Of Port Harcourt, Nigeria
[3] Department Of Computer Science Federal University Otuoke, Nigeria

ABSTRACT

One of the major challenges with most swarm intelligence algorithms in multimodal optimization is premature convergence. Search engines are using web spiders to crawl the web in order to collect copies of the web sites for their databases. These spiders usually use the technique of breadth first search which is non- guided or blind that depends on visiting all links of any web site one- by-one. We have studied various swarm intelligence algorithm with the search engine to enrich the literature. In this work, we develop a Web-based Crawler Using Bee Swarm Intelligent Algorithm. The methodology used is Structured System Analysis and Design Methodology(SSADM) in this approach. We implemented with PHP programming language and MySQL as database. The expected results show links into different module in a closed system, which appear to be more productive, fast and more flexible when compared with existing web crawlers. This work could be beneficial to the government, research communities and any other organisations that deals with information retrieval.

Keywords: Web Crawler, Information retrieval, Bee algorithm.

1.0 Introduction

Swarm Intelligence (SI) is a moderately new interdisciplinary field, which has increased gigantic prevalence nowadays. It is investigation of computational frameworks, which draw motivation from the aggregate knowledge straightforward specialists (like honey bees, ants and flying creatures). The most surely understood ideal models of swarm insight are Insect State Advancement which used ACO (Liao et al, 2012), Molecule Swarm Enhancement used PSO (Kirchmaier et al, 2012), Manufactured Honey bee Province used ABC (Karaboga et al, 2015), Stochastic Dissemination Hunt used SDS and bacterial rummaging calculation. Other nature-roused meta-heuristic calculations that proposed as augmentations of SI, for example, firefly calculation (Xin-She Yang, 2010), firecrackers calculation, wasp swarm calculation, Glowworm Swarm Enhancement (GSO), Gravitational Pursuit Calculation (GSA) et cetera.

An occupations of any web search tool is gathering pages likewise called creeping; web crawler is a product or program that uses the graphical structure of the web to move from page to page and add them to a neighborhood database. Web crawlers used to make a duplicate of all sent information by pages for later handling via web crawler, another definition for the web crawler (otherwise called robot) is a framework for downloading of the site pages, web crawlers utilized for assortment of purposes (Stoddard et al, 2012).

Numerous sorts of enhancement calculations are proposed to enhance the decent variety of the populace for counteracting untimely merging. Some of them are of swarms and groups in nature. Additionally, chasing and seeking conduct of predators are actualized by an ever increasing number of analysts and turned out to be a powerful technique. For example, the fundamental thought of Counterfeit Fish-Swarm Calculation that made used of AFSA (Xiaolei Li, 2013) is to mirror angle conduct, for example, preying, swarming, and following with neighborhood scan of individual fish for achieving the worldwide ideal. The Dim Wolf Streamlining agent that made use of GWO (Seyedali et al, 2014) calculation, mirrors the administration order and chasing system of dim frauds.

Numerous examinations about the swarm worldview (Christopher et al, 2010) have discovered that the gbest write focalizes rapidly on issue arrangements, however, has a soft spot for getting to be caught in the neighborhood optima, while lbest populaces can escape from nearby optima, as subpopulations investigate distinctive locales. In 2010 Kennedy hypothesized that heterogeneous populace structures, with a few subsets of the populace are firmly associated and others generally disconnected, could give the advantages of both lbest and gbest sociometries. Propelled by the heterogeneous populace structure, an adjusted dynamic virtual group is introduced in this paper with the point of quickening the early meeting rate and enhancing the worldwide hunting ability down FPO. Web crawler construct advancement in light of honey bee swarm knowledge calculation Dynamic virtual group, which was first displayed by the Dolphin Accomplice Enhancement (DPO) (Yang Shiqin et al, 2014) copies the chasing system of dolphins in nature. The execution of DPO with the

virtual group demonstrated the assessment of a few benchmark capacities.

Examinations demonstrate that it could exhibit the alluring execution.

In any case, the first structure of the dynamic group is not gainful for the FPO framework. Besides, the usage of dynamic virtual group in FPO caused a slight execution debasement. All things considered, the individual autonomous cognizance is debilitated by the group pioneer. Our work proposes a proper wheel topology rather than the first topology structure in DPO. In this paper, the adjusted structure of the dynamic virtual group as a suitable wheel topology is connected to FPO, which is a Dynamic Wellness Predator Enhancer (DFPO).

2.0 Related works

Tereshko et al, (2014): introduced an application for organic recreation using hereditary qualities calculations which nourishment source, a forager honey bee assesses a few properties related with the sustenance source, for example, its closeness to the hive, wealth of vitality, taste of its nectar, and the simplicity or trouble of separating this vitality. For the effortlessness, the nature of a sustenance source can be spoken to by just a single amount in spite of the fact that it relies upon different parameters specified before. Additionally, She conveys data about this particular source and offers it with different honey bees holding up in the hive. Notwithstanding their discovers, they likewise took a shot at Jobless foragers searches for a nourishment source to abuse. It can be either a scout who looks through the earth arbitrarily or a spectator who tries to discover a sustenance source by methods for the data given by the utilized honey bee yet they could not improve Counterfeit Honey bee Province (ABC) calculation.

Kuldeep et al, (2015): Presented, the Manufactured Honey bee Province (ABC) advancement calculation has connected by its counterfeit honey bee settlement was presented by Karabora in 2005. It was produced to tackle genuine parameter streamlining issue. BCO's (Honey bee Province Improvement) searching conduct is reproduced in manufactured honey bee state. This framework is depicting sorted out collaboration all around facilitated cooperation, work division, synchronous errand execution and well-sew

correspondence. In any case, they could not utilize fastidious honey bee framework.

Yonezawa and Kikuchi (1996), Seeley and Buhrman (1999), Schmickl et al. (2005) Lemmens (2006) all dealt with Natural reenactment utilizing fractional swarm smart yet the outcome were sufficiently bad in the earth of the honey bee.

Schmickl et al. (2005) assessed the heartiness of honey bees' rummaging conduct by utilizing a multiagent reenactment stage. They looked at the time-example of ecological changes influences the rummaging system and the productivity of the scavenging. They presumed that the aggregate searching technique of a bumble bee province is vigorous and versatile, and that its development enabled the settlement to discover ideal arrangements yet they couldn't utilize nonstop streamlining calculation to tackled the issue of bumble bee in its condition.

Lemmens (2006) investigated whether pheromone-based navigational calculations (enlivened by organic subterranean insect settlement conduct) are beated by non-pheromone-based navigational calculations (motivated by natural honey bee state conduct) in the undertaking of scavenging. The consequences of the investigations demonstrated that pheromone-based navigational calculations utilize less time per emphasis venture in little estimated universes,non-pheromone-based calculations are altogether quicker when finding and gathering sustenance and utilize less time ventures to finish the undertaking, and with developing world sizes, the non-pheromone-based calculation in the long run beats pheromone-construct calculations in light of a period for every time step measure. Regardless of every one of these benefits, it is said that non-pheromone-based calculations are less versatile than pheromone-based calculations.

3.0 Materials and methods

The Existing system

The current framework really influences utilization of Honey bee to swarm calculation to represent the Hive demonstrate. The calculation utilized by (Navrat et al, 2014) just inquiry through the web utilizing Meta-watchwords as it were. Each honey bee picks another source to take

after with likelihood equivalent to the quantity of honey bees moving for the source partitioned by the quantity of all moving honey bees (PSjf). With the contrary likelihood the honey bee remains in the assembly hall. On the off chance that the honey bee does not pick any source inside the predefined time (maximal time in assembly room, past which starvation would be fast approaching) she leaves the theater and goes to the dispatch space to pick the source from that point.

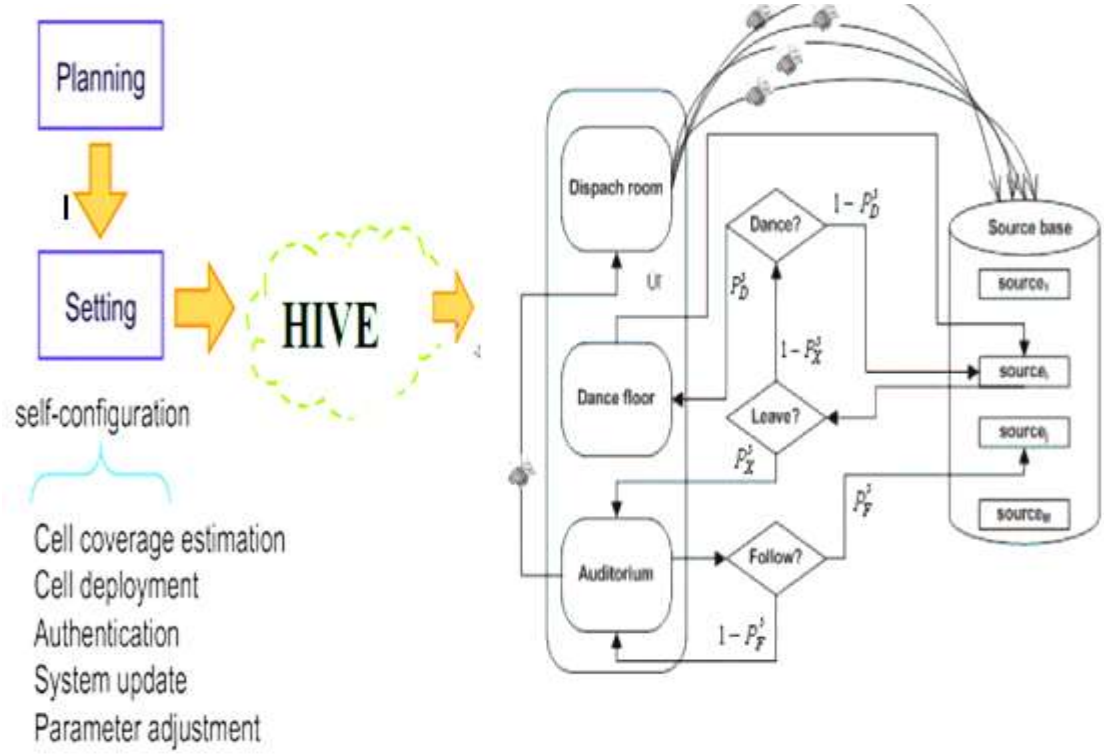

Figure 1.0: Architecture of Existing system (Navrat et al, 2014)

Drawbacks of Existing Framework includes

a. The model utilized as a part of the current framework can not characterize the conduct of the honey bee outside the hive. This all inclusive statement conveys a chance to adjust the conduct of the calculation to the particular needs of

the issue area without the need to alter the essential conduct of the hive.

b. Non-ideal: Since swarm frameworks are very excess and have no focal control, they have a tendency to be wasteful. The designation of assets is not productive, and duplication of exertion is constantly wild.

c. Uncontrollable: It is exceptionally hard to practice control over a swarm.

d. Unpredictable: The multifaceted nature of a swarm framework prompts unforeseeable outcomes.

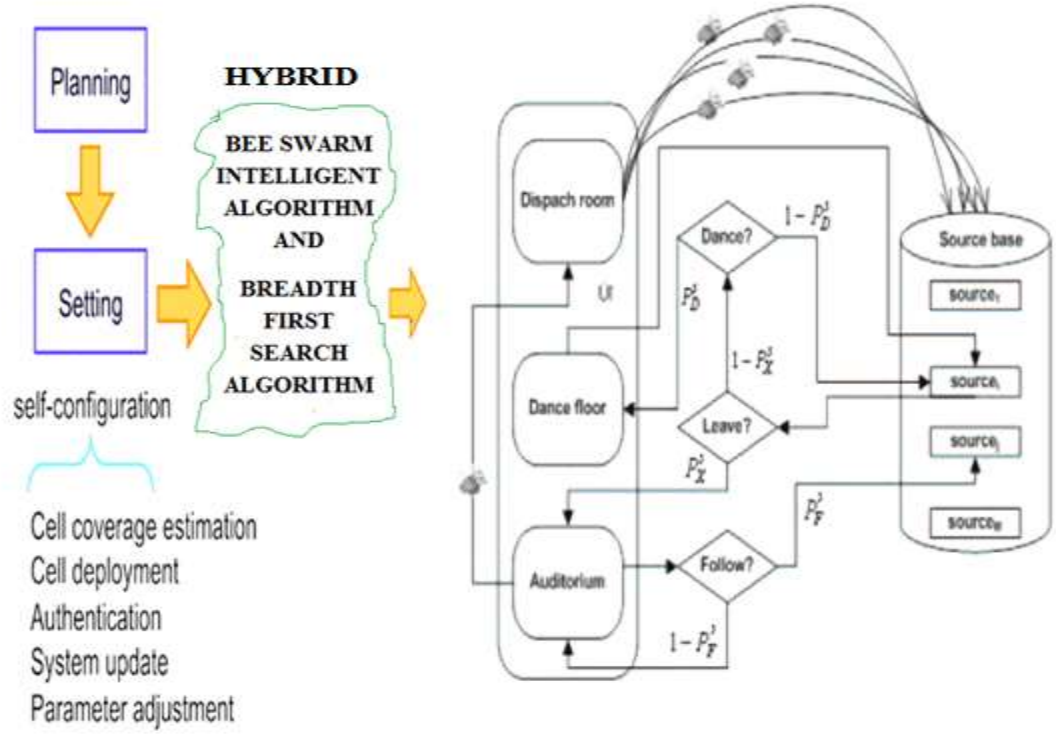

Figure 2.0: Architecture of the Proposed System

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 15
May 2018

The Hybrid Algorithm

The hybrid algorithm used in the proposed system comprises of Bee swarm and Breadth first algorithm, which tend to solve the crawling for information faster than the existing system. This hybrid algorithms crawl through the web using Normal search directory, meta-keywords and search keyword. The following algorithm illustrates the Hybrid algorithm in the proposed system functions and how it works:

a. Start
b. Initialize a crawling function by entering a seed URL to it.
c. Fetch the URL and save its URL keywords and page text in the database table called "Web data".
d. Find the URL's (links) to another pages.
e. Calculate the link relevancy using the following equation (page rank equation)
    i. PR(A)=(I-D)+D(PR(T1)/C(T1)+…+PR (Tn)/C(Tn))
f. For each link in the queue
g. If the link rank>0.7 then initialize a new function (bee) to crawl this link

Parameter Initialization

n = Number of employed bees

m = Number of onlooker bees (m>n)

Iteration: Maximum iteration number

αj: initial value of penalty parameter for jth agent

EC-Length: Length of ejection chain neighbourhood

Step: 1. Initialize employed bees with GRAH algorithm

σ i : ith employed bee in the population

Step: 2. Evaluate employed bees

Fitness Function (for minimization)

Step: 3. Repeat Cycle = 1

Number of Scout bees = 0,1*n

For each Employed Bee

Apply SHIFT Neighbourhood

If fit (Shift Neighbour)<fit(Employed Bee) then

Employed Bee = Shift Neighbour

Apply DOUBLESHIFT Neighbourhood

If fit (Double Shift Neighbour)<fit(Employed Bee) then

Employed Bee = Double Shift Neighbour

Determine probabilities by using fitness function

$$P_1 = \frac{\sum(1/ft)^{-1}}{ft}$$

(for minimization)

Calculate the number of onlooker bees which will be sent to food sources of employed bees, according to previously determined probabilities

Ni= Number of onlooker bees sent to ith sites = pi*m

Oij: jth onlooker bee of ith solution (j=1,…,Ni)

{Oi1, Oi2,…,OiNi}=EjectionChain(σ i)

Calculate fitness values for each onlooker bee

If the best fitness value of onlooker bees is better than the fitness value of employed bee, employed bee solution is replaced with this onlooker solution.

If (min (fit(Oij))<fit(σ i) then σ i = Oij

Best Solution

If fit (BestCycle-1)> Min(Fit(σ i))i=1..n then BestCycle=σ i
Else BestCycle= BestCycle-1
Until (i=n)

Step: 4. Scout bees
Initialize scout bees with GRAH algorithm

Step: 5. Cycle = Cycle+1
Until (cycle =Iteration)

Go to H

h. Successful search through Breadth first,
i. End for
j. End

Advantages of the Proposed System

The proposed system also uses Adaptive A* which is part of algorithm is the Best First Search algorithm.

1) It is use to search for the shortest paths repeatedly.
2) It uses its experience with earlier searches in the sequence to speed up the current A* algorithm.
3) Search and run faster than Repeated Forward A* algorithm.

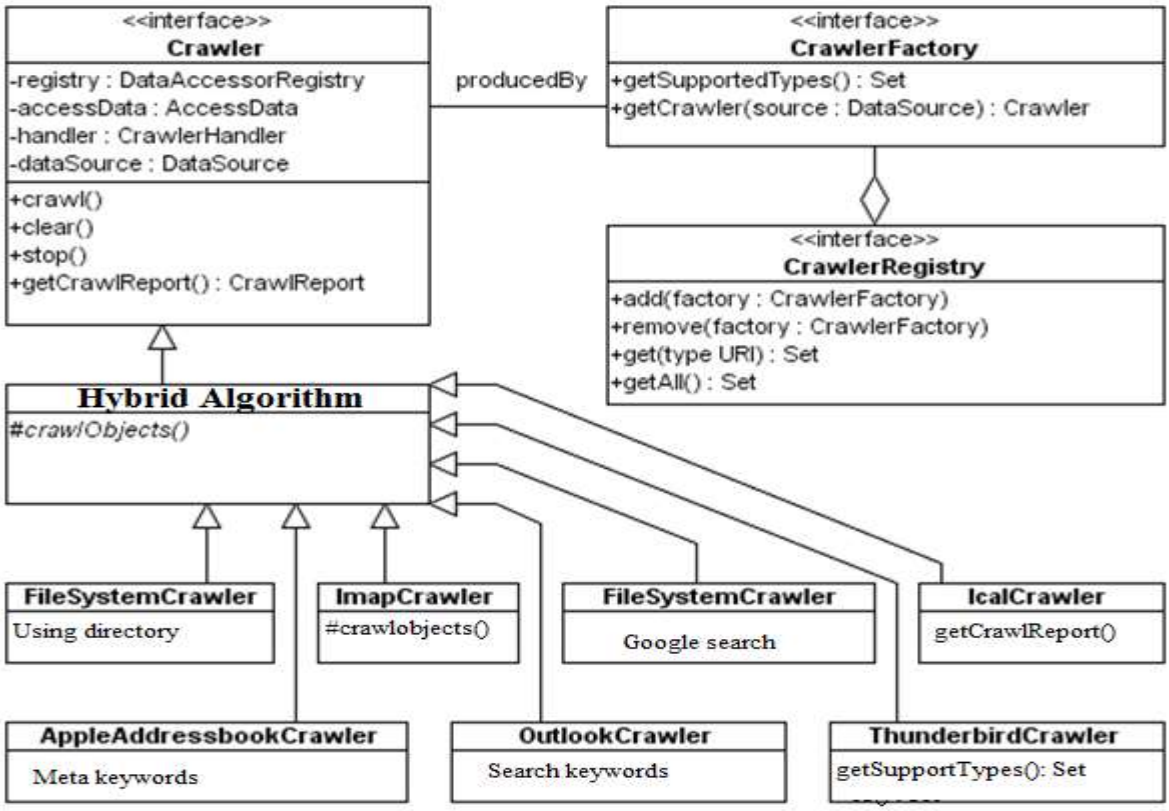

Figure 3.0: UML diagram for the Proposed system

4.0 Results and Discussion

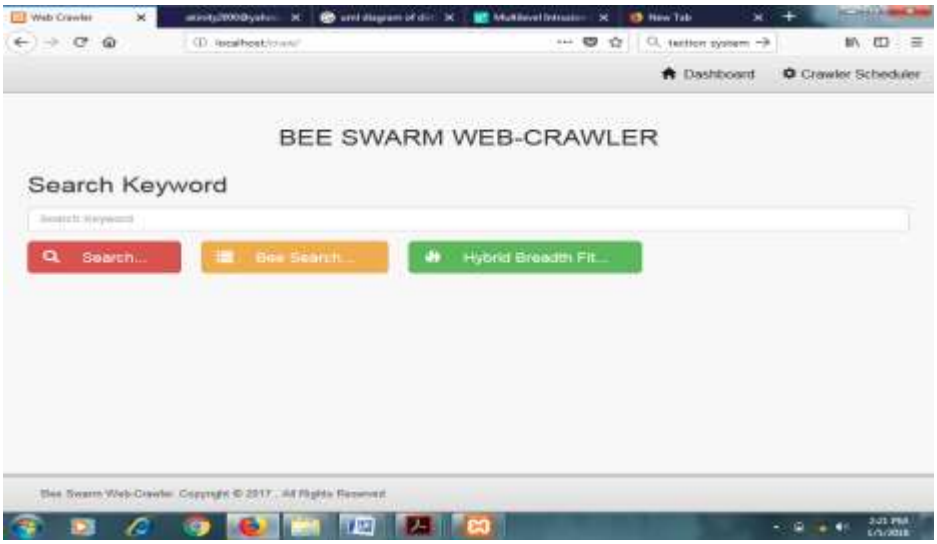

Figure 4.1(a): Search Result of Proposed System

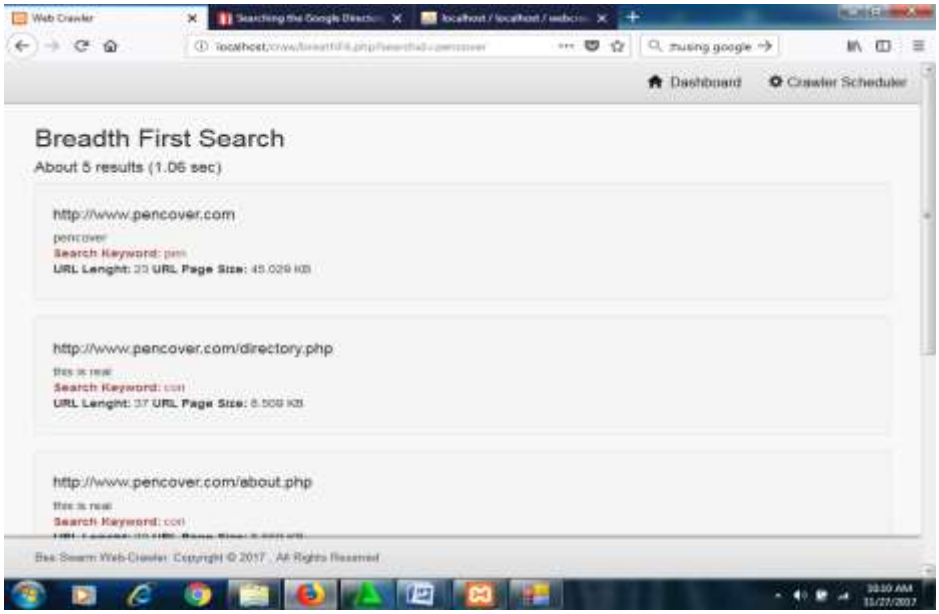

Figure 4.1(b): Search Result of Proposed System

Figure 4.1(a) is showing the interface of the software model while the figure 4.1(b) illustrated the used of the software to search and crawl for information.

Performance Evaluation

The result for selected tasks in comparing the Bee swarm algorithm and our Hybrid algorithm are based on software quality assurance measurement called METRICS. These sites have been crawled using the proposed bee algorithm and breadth first search and the result for crawling speed for each sit was as follows

Table 1.0: Results of the Proposed System

Website (keywords)	**Time of bee algorithm Crawling (Hive)**	**Time of Hybrid search crawling**
Pencover	**1.07sec**	**1.04sec**
Ojoma	**1.04**	**1.01sec**
Nairaland	**1.04sec**	**1.02sec**
Hair	**1.0sec**	**0.90sec**

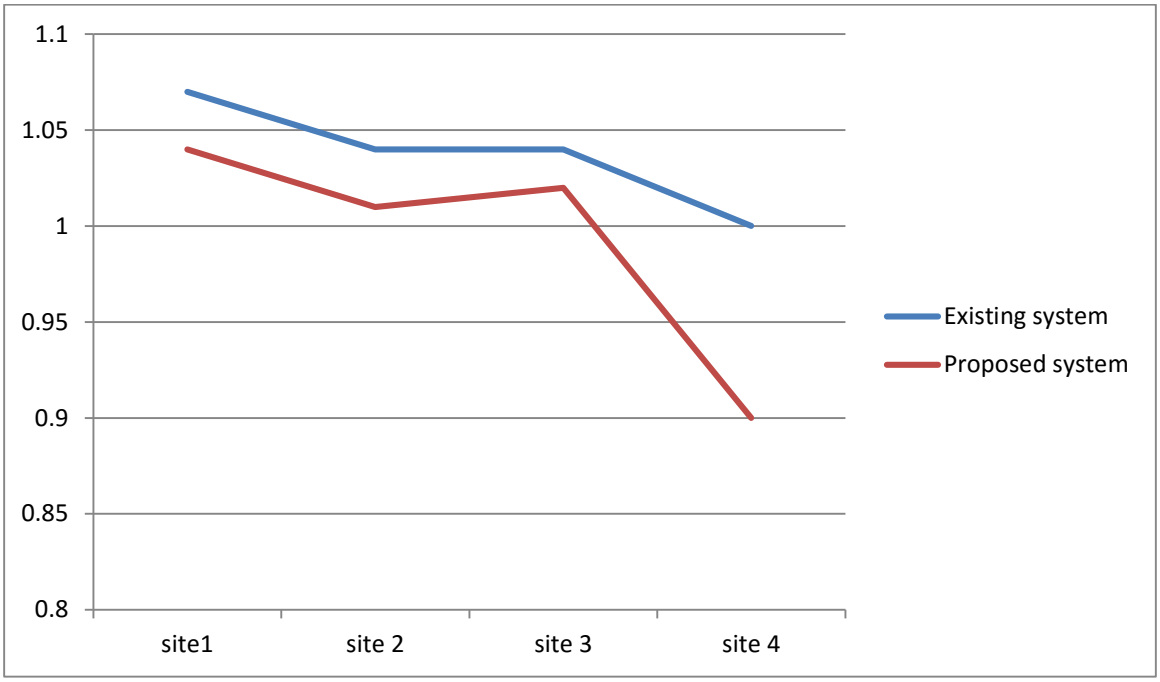

Figure 4.2: Graph illustrating and comparing Existing and Proposed system

In the obvious chart the reader can recognize that Hybrid algorithm crawling give a good result in time than the Hive crawling. Now, the second test is the test of crawled pages relevancy by crawling the web for specific subjects, so the user can input to the system terms like (Keywords: free, computer, books, and download) with spaces between each term, then press on "crawl the web from Google directory", this system will crawl a sample of 5 sites , now repeat the operation by using the last button "crawl using Hybrid (breadth first search and hive).

5.0 Conclusion

This research reach to a good result when using the hybridized swarm algorithm for crawling a single web site in speed, and also we have good results in crawling the web

using keywords for a group of sites in sped and relevancy of the crawled sites to the desired topics, so the a swarm intelligence can be viewed as a good improvement in web crawling area.

Contribution to knowledge

1. We developed a Web-based Crawler Using Bee Swarm Intelligent Algorithm
2. We have designed a proposed model using hybrid algorithm that gives room for relatively fast search of information within a short period of time.
3. It can also be used to test the genuity of a real website.

6.0 Recommendations

The work carried out in this research is recommended to organizations both private and public (e.g. surveillance, Universities, E-government system etc.) that are looking for a better and faster way to search for information for proper used.

7.0 References

Akay, B (2013): A study on particle swarm optimization and artificial bee colony algorithms for multilevel thresholding, Appl. Soft Computing. 13) 3066–3091.

Amir R. and Amir S ,(2012): Biology of Spiders, 198 Madison Ave. NY, New York, 10016: Oxford University Press,

Bank K. M. and Passino H, (2013): Biomimicry of bacterial foraging for distributed optimization and control, IEEE Control Syst. Mag. 22 (3), 52–67.

Bonabeau, M. Dorigo, G. and Theraulaz, E(2014): "Swarm Intelligence: From Natural to Artificial Systems", New York, NY: Oxford University Press,.

Christopher O and Mark N, (2010): "web crawling", yahoo research, "An Idea Based on Honey Bee Swarm For Numerical Optimization" (PDF).

Christopher oilstone and Mark najork,(2010): "web crawling", yahoo research.

Chu W, X. Gao X, and Sorooshian S (2011): Handling boundary con- straints for particle swarm optimization in high-dimensional search space, Inform. Sci. 181 (20) 4569–4581.

Das, P. N. Suganthan S, (2011): Differential evolution: A survey of the sate-of-the-art, IEEE Trans. Evol. Comput. 15 (1) 4–31

De Castro L.N. and F.J. Von Zuben, (2014): "Artificial Immune Systems. Part I. Basic Theory And Applications", Technical Report No. Rt Dca 01/99, Feec/Unicamp, Brazil,.

Dorigo M and Birattori K. M. (2014):Ant colony optimization for continuous domains, European Journal of Operational Research 185 (3), 1155–1173.

Drias H., Sadeg S., Yahi S.,(2005): Cooperative Bees Swarm for Solving the Maximum Weighted Satisfiability Problem, *IWAAN International Work Conference on Artificial and Natural Neural Networks*, Barcelona, Spain, 318-325.

Farhoodnea, A. Mohamed, H. Shareef, H. and Zayandehroodi M.(2014): Optimum placement of active power conditioner in distribu- tion systems using improved discrete firefly algorithm for power quality enhancement, Appl. Soft Comput. (23)249–258.

Gianni, di carro,(2015) "introduction to swarm intelligence", http://www.idsia.ch/~gianni/lectures/swarm-part-1.pdf

James, . R. L. and shareef H (2015) :"The Evolution of the Organization of Work in Social Insects" Monit. Zool. Ital. 20, 267-287.

Navrat, P. Drias H., Sadeg S., Yahi S (2008): Web Search Engine Working as a Bee Hive. *Web Intelligence and Agent Systems*, 6:441 – 452,

James K, (2010): Particle swarm optimization, In Encyclopedia of Machine Learning,760–766.

karabog D (2005):A., "a comparative study of artificial bee colony algorithm", ELSEVIER, www.elsevier.com/locate/amc, 2009.

Karaboga H. and Dervis G. (2015): "An Idea Based on Honey Bee Swarm For Numerical Optimization" (PDF).

Kennedy, R. C. Eberhart J (2013): "Particle swarm optimization", In Proceedings of the IEEE International Conference on Neural Networks" (4) 1942–1948.

Kirchmaier, S. Hawe, K. Diepold S (2013), A swarm intelligence in- spired algorithm for contour detection in images, Appl. Soft Comput. 13(4) 3118–3129.

Lam, V. O. K. Li, J. J. Q. Yu A. Y. S. (2013) Real-coded chemical reaction optimization, IEEE Trans. Evol. Comput. 16 (3)339–353.

Liao T., Molina D., Stutzle T., Oca M. and Dorigo M., (2012): An ACO algorithm benchmarked on the bbob noiseless function testbed, in: Proc. 14th Int. Conf. GECCO, Philadelphia, U.S., 221-228.

Mallipeddi, R S. Mallipeddi, P. N. Suganthan, M. Tasge- tiren, F.(2011): Differential evolution algorithm with ensemble of param- eters and mutation strategies, Appl. Soft Comput. (11) 1679–1696.

Mark Fleischer, (2012) "foundation of swarm intelligence", Institute of system research, university of Maryland.

Mrco dorigo and Mauro birattari, (2014) "swarm intelligence",http://www.scholarpedia.org/article/swarm_intelligence,2007

Omkar, J. Senthilnath, R. Khandelwal, G. N. Naik, S. Gopalakrishnan,(2011): Artificial bee colony (abc) for multi-objective design optimization of composite structures, Appl. Soft Comput. (11), 489–499.

Oster G, and Wilson E. O.(2015): "Castes and Ecology in the Social Insects", Princeton, NJ: Princeton University Press.

Parpinelli R.S, Lopes H.S, (2011): New inspirations in swarm intelligence: a survey, Int. J. Bio-Inspired Computation 3 (1), 1-16.

Pham D.T., Kog E., Ghanbarzadeh A., Otri S., Rahim S., Zaidi M., (2006a) The Bees Algorithm – A Novel Tool for Complex Optimisation Problems, *IPROMS 2006 Proceeding 2nd International Virtual Conference on Intelligent Production Machines and Systems*, Oxford, Elsevier.

Pham D.T., Otri S., Ghanbarzadeh A., Kog E.,(2006b) Application of the Bees Algorithm to the Training of Learning Vector Quantisation Networks for Control Chart Pattern Recognition, *ICTTA'06 Information and Communication Technologies*, 1624-1629.

Qin A. K. X. Li, (2013): Differential evolution on the CEC-2013 single-objective continuous optimization testbed, in: Proc. IEEE Congress on Evolutionary Computation (CEC), Cancun, Mexico, 1099–1106.

Sandeep S. (2014):"web-crawling approaches in search ",Thapar university, Patiala, India.

Schaber, S. N. Gorb, F. G. and Barth C. F. (2012): Force transformation in spider strain sensors: White light interferometry., J. Royal Society Interface 9 (71), 1254–126

Seeley T. D. and Visscher P.K.(2015): "Assessing the benefits of cooperation in honeybee foraging: search costs, forage quality, and competitive ability", Behav. Ecol. Sociobiol., 22: 229-237.

Seyedali M., Seyed, M. M., and Andrew L., (2014): Grey wolf optimizer, Adv. Eng. Software. 69, 46–61

Shiqin Y. and Yuji S. (2014): Fitness predator optimizer to avoid premature convergence for multimodal problems, In Systems, Man and Cybernetics, IEEE International Conference IEEE 258–263

Stoddard P. K., Salazar V. L., (2012): Energetic cost of communication, The Journal of Experimental Biology 2(14), 200-205.

Swagatam D., Arijit, B., Sambarta D., and Ajith A. (2016): Bacterial foraging optimization algorithm: theoretical foundations, analysis, and applications, In Foundations of Computational Intelligence (3),23–55

Tereshko,V (2011): "Reaction-diffusion model of a honeybee colony's foraging behaviour, M. Schoenauer, et al, Eds., Parallel Problem Solving from Nature VI", Lecture Notes in Computer Science, Springer-Verlag: Berlin. 19(17), 807-816.

Vesterstrøm, J. Riget J. (2012): Particle Swarms Extensions for improved local, multi-modal, and dynamic search in numerical optimization, MSc. Thesis, May

Wenping Z., Younlong Z., Hanning C. and Zhu Z. (2010): "cooperative approaches to artificial bee colony algorithm" , international conference on computer application and system modeling, 23-45.

Xiaolei Li,(2013): A new intelligent optimization-artificial fish swarm algorithm, Doctor thesis,34-78.

Xin-She Y., (2010): A new metaheuristic bat-inspired algorithm, In Nature inspired cooperative strategies for optimization. 65–74

Yang S., Jiang J., and Yan G., (2014): A dolphin partner optimization, In Proceedings of the 2009 WRI Global Congress on Intelligent Systems 1(9), 124–128

Yim M., Zhang Y., and Duff .D. (2012): Modular robots, IEEE Spectrum 39 (2), 30-34.

Yuhui S. (2011): Brain storm optimization algorithm, In Advances in Swarm Intelligence, Springer, 303–309