Information Retrieval using Soft Computing: An Overview

Md. Abu Kausar, Md. Nasar, Sanjeev Kumar Singh

Abstract— Information retrieval is a broad research area generally in the World Wide Web (WWW). A huge Repository of web pages, images in the World Wide Web causes result which search engines provide, it holds a lot of irrelevant information. Thus, finding the information which user needs has become hard and complicated, so that proper retrieval of information has become more important. Facing with these challenges, using of new techniques such as genetic algorithm, differential evaluation, artificial neural network, ant colony algorithm can be a suitable method to solve these problems. In this paper we will investigate the use of these techniques in information retrieval.

Index Terms— Search Engine, Genetic Algorithm, Information Retrieval, Differential Evolution, Neural Network, Ant Colony Algorithm, Web Crawler.

———————————— ◆ ————————————

1 INTRODUCTION

In the past three decade, the area of information retrieval (IR) has developed well beyond its primary goals of indexing text and searching for useful documents in a collection. Currently, research in IR includes document classification, modeling and categorization, systems architecture, data visualization, user interfaces, filtering, languages, etc. IR was seen as a broad area of interest generally to librarians and information retrieval experts. Such a tendentious idea overcomes for many years, although the rapid distribution, among users of modern computers of Information Retrieval tools for multimedia and hypertext applications.

The Web is becoming a universal repository of text, document, images and music etc. in a large scale never seen before. Its success is based on the idea of a standard user interface which is always the same no matter what computational environment is used to run the user interface. As a result, the user is protected from details of communication protocols, machine location, and operating systems. Further, any user can create his own Web documents and make them point to any other Web documents without restrictions. This is a key aspect because it turns the Web into a new publishing medium accessible to everybody. As an instant consequence, any user can push his/her personal plan with little effort and almost at no cost. This universe without limits has attracted tremendous attention from millions of people everywhere since the very beginning. Furthermore, it is causing a revolution in the way people use computers and perform their daily tasks. For instance, online shopping and home banking through internet are becoming very popular and have generated several hundred million dollars in revenues.

Despite too much achievement, the WWW has initiated latest problems of its own. Finding valuable information on the Web is frequently a tedious and difficult task. For example, to satisfy his information need, the user may navigate the space of Web links searching for information of interest. However, since the hyperspace is huge and almost unknown, such a navigation task is usually inefficient. For new users, the problem becomes harder, which might entirely frustrate all their efforts. The main difficulty is the absence of a well defined underlying data model for the Web, which implies that information definition and structure is frequently of low quality. These complexities have attracted improved interest in Information Retrieval and its techniques as promising resolutions. As a result the Information Retrieval has achieved a place with other technologies. The detail about this is also discussed by [46].

Now days it has become an important part of human life to use Internet to gain access the information from WWW. The current population of the world is about 7.017 billion out of which 2.40 billion people (34.3%) use Internet [3] (see Figure 1). From .36 billion in 2000, the amount of Internet users has increased to 2.40 billion in 2012 i.e., an increase of 566.4% from 2000 to 2012. In Asia out of 3.92 billion people, 1.076 billion (i.e.27.5%) use Internet, whereas in India out of 1.2 billion, .137 billion (11.4%) use Internet. Same growth rate is expected in near future too and it is not far away when anyone will start believing that the life is incomplete without Internet. Figure 1: illustrates Internet Users in the World Regions.

————————————————

- *Md. Abu Kausar is currently pursuing Ph.D in Computer Science in Jaipur National University, Jaipur, India. E-mail: kausar4u@gmail.com*
- *Md. Nasar is currently working with Dept. of Computer Application, Galgotias University, Gr. Noida, India. E-mail: nasar31786@gmail.com*
- *Dr. Sanjeev Kumar Singh is currently workingwith Dept. of Mathematics Galgotias University, Gr. Noida, India. E-mail : sksingh8@gmail.com*
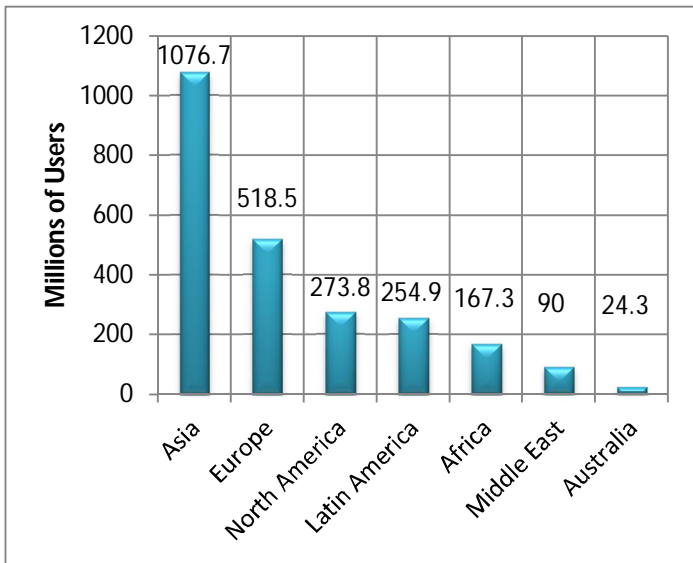
Fig. 1. Internet Users in the World Regions (Source: http://www.internetworldstats.com accessed on Feb 15, 2013)

2 SEARCH ENGINE

Beginning in 1990, World Wide Web has grown exponentially in size. As of today, it is estimated that it contains about 55 billion publicly index able web documents [46] spread all over the world on thousands of servers. It is not easy to search information from such a vast collection of web documents available on WWW. As the web page around the world is increasing day by day, the need of search engines has also emerged. We explain the basic components of any basic search engine along with its working.

2.1 Basic Web Search Engine

The abundant content of the World-Wide Web is useful to millions. Some people simply browse the Web through entry points such as Google, Yahoo etc. But many information seekers use a search engine to begin their Web activity, users submit a query, usually a list of keywords, and receive a list of Web pages that may be relevant to the query, normally pages that contain the keywords. By Search Engine in relation to the Web, we are usually referring to the actual search form that searches through databases of HTML documents.

2.2 Types of Search Engine

There are basically three types of search engines:
1. Powered by human submissions
2. Powered by robots (called crawlers, ants or spiders)
3. And those that are a hybrid of the two.

Two main types of search engines are explained briefly below.

2.2.1 Human-powered Search Engine

Such search engines rely on humans to submit information that is subsequently indexed and catalogued. Simply information that is submitted is placed into the index. This type of search engines are rarely used at large scale. But these are use-ful in the organizations where small scale of data is dealt with.

2.2.2 Crawler Based Search Engine

Such search engines use automated software agents (called crawlers) that visit a Website, read the information on the actual site, read the web site's meta tags and also follow the links that the web site connects to performing indexing on all linked sites as well. The crawler returns all information back to a central repository, where the data is indexed. The crawler periodically returns to the web sites to check for any information that has been changed.

2.3 Structure and Working of Search Engine

The fundamental structure of crawler based search engine is shown in figure 2. Thus the main steps in any search engine are
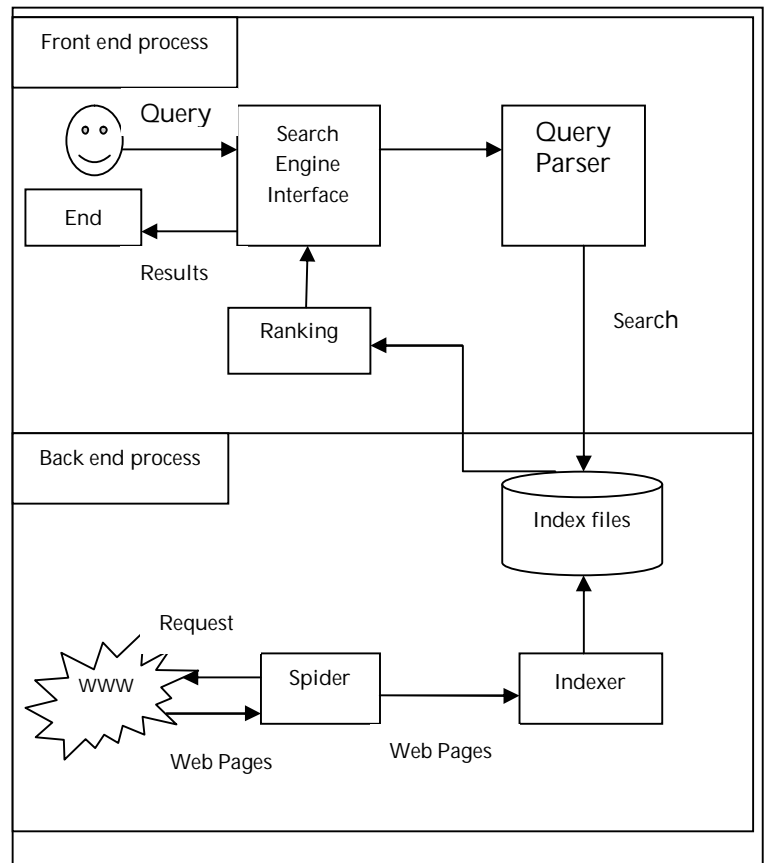

Fig. 2. Working steps of search engine

Crawling

Every search engine relies on a crawler to provide the grist for its operation. This operation is performed by special software, called Crawlers. Web crawler is a program/software or programmed script that browses the WWW in a systematic, automated manner on the search engine's behalf. The programs are given a starting set of URLs called seed URL, whose pages they retrieve from the Web. The web crawler extracts URLs appearing in the retrieved pages, and provides this information to the crawler control module. This module determines which links to visit next, and feeds the links to visit back to the crawlers.

Maintaining Repository

All the data of the search engine is stored in a repository as shown in the figure 2. All the searching is performed through that database and it needs to be updated frequently. During a crawling process, and after completing crawling process, search engines must store all the new useful pages that they have retrieved from the WWW.

Indexing

Once the web pages are stored in the repository, the next job of search engine is to make index of stored data. The indexer module extracts all the words from every web page, and records the URL where each word occurred. The result is a usually very large that can provide all the URLs that point to web pages where a given word occurs.

Querying

This module deals with the user queries. The responsibility of query engine module is for receiving and filling search requests from users. The search engine relies deeply on the indexes, and sometimes on the page repository.

Ranking

Since the user query results in a large number of results, it is the work of the search engine to display the most suitable results to the user. To do this efficient searching, the ranking of the results are performed. The ranking module has the task of sorting the results such that results near the top are the most likely ones to be what the user is looking for. Once the ranking is done by the Ranking module, the final results are displayed to the user.

3   GENETIC ALGORITHM

A genetic algorithm is a search procedure inspired by principles from natural selection and genetics. It is often used as an optimization method to solve problems where little is known about the objective function. The operation of the genetic algorithm is quite simple. It starts with a population of random individuals, each corresponding to a particular candidate solution to the problem to be solved. Then, the best individuals survive, mate, and create offspring, originating a new population of individuals. This process is repeated a number of times, and typically leads to better and better individuals.

Genetic Algorithms works with a set of individuals, representing possible solutions of the task. The selection principle is applied by using a criterion, giving an evaluation for the individual with respect to the desired solution. The best-suited individuals create the next generation.
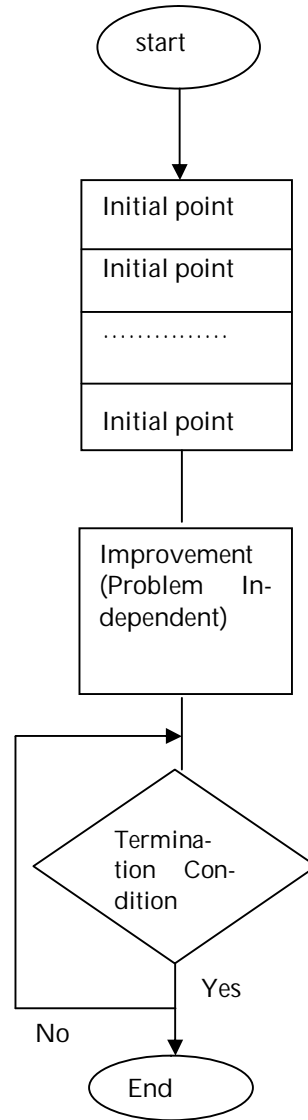

Fig. 3. Working steps of Genetic Algrithm

4   ARTIFICIAL NEURAL NETWORK

An Artificial Neural Network (ANN) is an information processing model that is inspired by the biological nervous systems, such as brain process information. ANN, like learns by example. Neural networks, with their extraordinary ability to obtain meaning from complicated or imprecise data, can be helpful to extract patterns and detect trends that are very complex to be noticed by either human beings or other computational techniques. A trained neural network will be thought of as an "expert" in the group of information it has been specified to analyze.

An artificial neural network (ANN) engages a network of simple processing elements artificial neurons, which will exhibit complex global behavior, resolved by the connections

between the processing elements and element parameters. In a neural network model, nodes, which can be called neurons, units or processing element are connected together to form network of nodes. ANN, often called a neural network (NN), is an interconnected group of artificial neurons that uses a computational model for information processing based on a connectionist approach. An adaptive ANN changes its structure based on information that flows through the network.

5 ANT COLONY ALGORITHM

The Ant Algorithm (AA) is born in modern years and derived from the natural bionic algorithm. AA was firstly proposed by M. Dorigo [33, 34] and the major idea of it is to search the optimal solution by using the information transmission way of the ants, This is called the ant colony optimization (ACO) algorithm. And for the apply of the concept of artificial ants, it is also called the ant system (AS).

The advantages of such an algorithm is:

1) The principle is called positive feedback mechanism or enhanced learning system, it is constantly updated by the pheromone to eventually converge to the optimal path
2) It is a distributed optimization method, not only for the current serial computers, but also for the future of parallel computer
3) It belongs to general purpose stochastic optimization methods, but it is by no means a simple simulation of real ants, and it is melt into the human intelligence
4) It is a global optimization method, not only for solving single objective optimization problems, but also can be used for multi-objective optimization problem.

Ants have the ability to find the nearest path without any prompting from their nest to food source. And they also can change with environment to search for new paths, create new choices adaptively. The main reason is that ants can release a special kind of secretion (pheromone) in the search process for food. And this special matter can volatile. The probability of path choice for the later ants is proportional to the strength of this kind of matter on the paths. So, when more and more ants choose the same path, the more pheromone will be left to increase the choice probability.

6 DEFRENTIAL EVOLUTION

Differential evolution (DE), proposed by Storn and Price [45], is a simple powerful population based stochastic search technique for solving global optimization problems. DE has been used successfully in various fields such as communication [43], pattern recognition [45], and mechanical engineering [44], to optimize non convex, non-differentiable and multi modal objective functions. DE has many striking properties compared to other evolutionary algorithms such as, implementation ease, the small number of control parameters, fast convergence rate, and robust performance. DE has only a few control variables which remain fixed throughout the optimization process, which makes it easy to implement. Moreover, DE can be implemented in a parallel processing framework, which

enables it to process a large number of training instances efficiently. This property of DE makes it an ideal candidate for the current task of learning a ranking function for information retrieval, where we must optimize non convex objective functions.
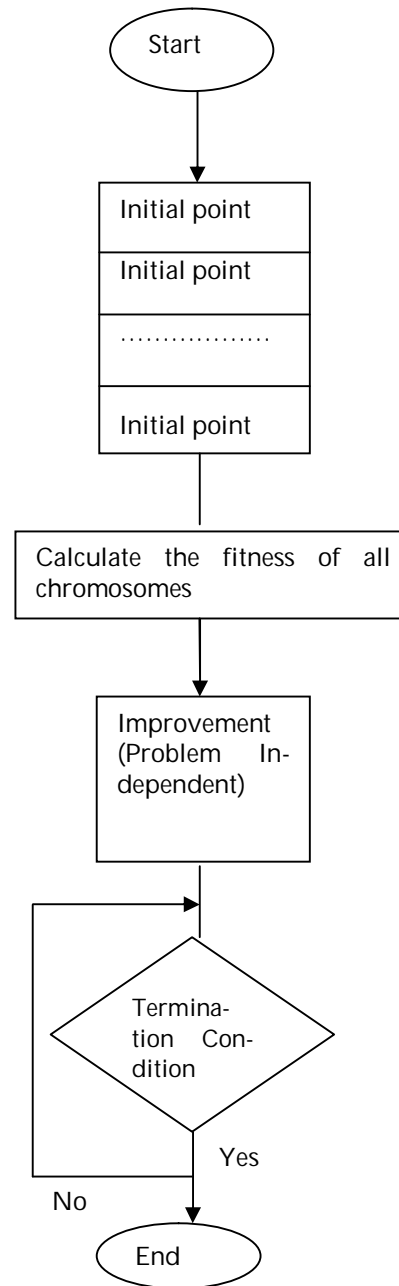

Fig. 4. Working steps of Diferential Evolution

7 LITERATURE REVIEW

Bangorn Klabbankoh and Ouen Pinngern [4] analyzed vector space model to boost information retrieval efficiency. In vector space model, IR is based on the similarity measurement between query and documents. Documents with high similarity to query are judge more relevant to the query and will be retrieving first. Testing result will show that information retrieval with 0.8 cross-

over probability and 0.01 mutation probability provide the maximum precision while 0.8 crossover probability and 0.3 mutation probability provide the maximum recall. The information retrieval efficiency measures from recall and precision.

Recall is defined as the proportion of relevant document retrieved.

Recall= (Number of documents retrieved and relevant) / (Total relevant in collection)

Precision= (Number of documents retrieved and relevant)/ (Total retrieved)

A tested database consisted of 345 documents taken from student's projects. Beginning experiment indicated that precision and recall are invert. To use which parameters depends on the appropriateness that what would user like to retrieve for. In the case of high precision documents prefer, the parameters may be high crossover probability and low mutation probability. While in the case of additional relevant documents (high recall) prefer the parameters may be high mutation probability and lower crossover probability. From beginning experiment specified that we can use GA's in information retrieval. The Work by Ahmed A. A. et. al. [5] developed a new fitness function for estimated information retrieval which is very quick and very flexible, than cosine similarity fitness function.

M. Koorangi and K. Zamanifar [6] analyzed the problems of current web search engines, and the need for a new design is necessary. Novel ideas to improve present web search engines are discussed, and then an adaptive methods for web meta search engines with a multi agent particularly the mobile agents is presented to make search engines work more proficiently. In this method, the assistance between stationary and mobile agents is used to make more efficiency. The meta-search engine presents the user needed documents based on the multi stage mechanism. The combine of the results got from the search engines in the network is done in parallel. In another work, Abdelmgeid A. Aly [7] discussed an adaptive method using genetic algorithm to change user's queries, based on relevance judgments. This algorithm is personalized for the three well-known documents collections (CISI, NLP and CACM). This method is shown to be appropriate for large text collections, where more appropriate documents are presented to users in the genetic modification. Alin Mihaila et. al. [8] Studied Text segmentation is an important problem in information retrieval and summarization. The segmentation process tries to split a text into thematic clusters (segments) in such a way that every cluster has a maximum cohesion and the contiguous clusters are connected as little as possible. Ziqiang Wang et. al. [9] presented Memetic algorithm which combines evolutionary algorithms with the intensification power of a local search, and has a pragmatic perspective for better effects than GA. As such Memetic algorithm, a local optimizer is applied to each offspring before it is inserted into the population in order to make it towards optimum and then GA platform as a means to accomplish global exploration within a population. Memetic algorithm is based on a vector space model in which both documents and queries are represented as vectors. The goal of MA is to find an optimal set of documents which best match the user's

need by exploring different regions of the document space simultaneously. The system ranks the documents according to the degree of similarity between the documents and the query vector. The higher the value of the similarity measure is, the closer to the query vector the document is. If the value of the similarity measure is sufficiently high, the document will be retrieved. The Memetic algorithm tries to involve, generation by generation, a population of queries towards those improving the result of the system. Author also compare the number of relevant document retrieved using MA, PSO and GA. Comparison of relevant document gives the number of relevant document retrieved at each iteration of the three optimization algorithm. Rong LI et. al. [11] presented Hidden Markov Model (HMM) is easy to establish, does not need large scale sample set and has good adaptability and higher precision. When extracting text information based on HMM, Maximum Likelihood (ML) algorithm for marked training sample set or Baum-Welch algorithm for unmarked training sample set is adopted generally to obtain HMM parameters. ML algorithm is a kind of local searching algorithm and Baum-Welch algorithm is one kind of concrete implementation of Expectation Maximum (EM) algorithm. GA-HMM hybrid model has been applied successfully in speech recognition; however its application in text information extraction has not been seen. An improved hybrid algorithm for text information extraction is proposed to optimize HMM parameters by using GA. Compared with the traditional training algorithm, GA has obvious superiority of seeking global optimum. Through the improvement on traditional GA and combination with text information characteristic, a hybrid algorithm for text information extraction based on GA and HMM is proposed. In HMM training process, the hybrid algorithm uses GA to seek the optimal solution. Loia and Luengo [10] presented an evolutionary approach useful to automatically construct a catalogue as well as to perform the classification of web documents. The proposal faces the two fundamental problems of web clustering the high dimensionality of the feature space and the knowledge of the entire document. The first problem is tackled with genetic computation while the authors perform a clustering based on the analysis of context in order to face the second one. The genome is defined as a tree based structure and two different evaluation functions are used. As genetic operators, the one-point crossover and five different mutation operators are defined. On the concept of Text Information Extraction based on Genetic Algorithm and Hidden Markov Model. Habiba et. al. [12] introduced hybrid Genetic Algorithm which shows that indeed for large scale collection, heuristic search techniques outperform the conventional approaches in addressing retrieval. Author proposed two evolutionary approaches have been designed and developed for information retrieval. The first one, namely GA-IR is a genetic algorithm and the second is an improved version towards a memetic algorithm called MA-IR. The aim of proposed study is the adaptation of heuristic search technique to IR and their comparison with classical approaches. Authors also concluded that both GA-IR and MA-IR are more suitable for large scale information retrieval than classical IR method and that MA-IR outperforms GA-IR. Lourdes Araujo and Joaquin Perez-Iglesias [13] studied Training a Classifier for the selection of Good Query Expansion Terms with a Genetic Algorithm. Authors developed a classifier which has been trained for differenti-

ating good expansion terms. The identification of good terms to train the classifier has been achieved with a genetic algorithm whose fitness function is based on user's relevance judgments on a set of documents.. It should be noted that the GA was very simple and does not includes any re-weighting for terms, i.e. only a Boolean representation was applied in order to model queries and terms. Pratibha Bajpai and Manoj Kumar [14] discussed global optimization and discussed how genetic algorithm can be used to achieve global optimization and demonstrate the concept with the help of Rastrigin's function. The objective of global optimization is to find the "best possible" solution in nonlinear decision models that frequently have a number of sub-optimal solutions. The genetic algorithm solves optimization problems. It helps to solve unconstrained, bound constrained and general optimization problems, and it does not require the functions to be differentiable or continuous. A. S. Siva Sathya and B. Philomina Simon [15] proposed document crawler which is used for collecting and extracting information from the documents accessible from online databases and other databases. The proposed information retrieval system is a two stage approach that uses genetic algorithm to obtain the set of best combination of terms in the first stage. Second stage uses the output which is obtained from the first stage to retrieve more relevant results. Thus a novel two stage approach to document retrieval using Genetic Algorithm has been proposed. The proposed information retrieval system is more efficient within a specific domain as it retrieves more relevant results. This has been verified using the evaluation measures, precision and recall. More recently, clustering will be used for helping the user in browsing a group of documents or in organizing the results returned by search engines [16]. In [17] the authors discussed a novel method of combining the clustering and genetic optimization in improving the retrieval of search engine results in diverse settings it is possible to design search methods that will operate on a thematic database of web pages that will refer to a common knowledge or to specific sets of users. They will consider such premises to design and develop a search technique that will deploy data mining and optimization techniques to give a more significant and restricted set of pages as the final result of a user query. They will accept a vectorization method that is based on search context and user profile to apply clustering methods that are then refined by genetic algorithm. As discussed in [16], the application of clustering in information retrieval (IR) is based typically on the cluster hypothesis. Numerous researchers have exposed that the cluster hypothesis also grasps in a retrieved set of documents, but they do not study how the clustering structure may help a user in finding relevant results more rapidly. Meta heuristics and more precisely, genetic algorithms have been implemented in information retrieval (IR) by numerous researchers and the results shows that these algorithms will be efficient. Gordon [18] discussed a genetic algorithm (GA) based approach to improve indexing of documents. In this approach, the initial population is generated by a collection of documents judged relevant by a user, which is then developed through generations and converges to an optimal population with a set of keywords which best explain the documents In [19] author adopted a similar concept to document clustering, where a genetic algorithm is used to adapt topic descriptions so that documents become more efficient in matching relevant queries. In

[23] Eugene Agichtein et. al. discussed that incorporating users activities data will significantly improve ordering of top results in real web search. They observed the alternatives for incorporating feedback into the ranking process and investigate the contributions of user feedback compared to different common web search features. In [20] the authors apply genetic algorithm (GA) in information retrieval (IR) in order to improve search queries that produce improved results according to user's choice. In [22] Zhongzhi Shi et. al. studied the existing methodology for Web mining, which is moving the WWW toward a more helpful environment in which users can quickly and easily find the information they needed. [21] Al-Dallal et. al. proposed a text mining approach for web document retrieval that applies the tag information of HTML documents that GA is applied to find important documents. In [24] YaJun Du et. al. discussed an intelligent model and it is a implementation of search engine that the process of searching information on Internet is similar as book search. Author proposed that Search Engines take on the five intelligence behaviors corresponding five parts intelligence of human kind. They divided the process of information searching of search engine into four stages classifying Web page, confirming a capacity of information searching, crawling Web pages, and filtrating the result Web pages. Neural networks computing, in particular, seems to fit well with conventional retrieval models such as the vector space model [35] and the probabilistic model [36]. In AIR [38] He developed a three-layer neural network of authors. Index terms, and web documents. The system will use relevance feedback from its users to modify its representation of authors, index terms, and web documents over time. The result was a representation of the consensual meaning of keywords and documents shared by some group of users. One of his main contributions is the use of a change correlation learning rule. The learning process created many novel connections between web documents and index terms. Doszkocs et al. [37] provided an excellent overview of the use of connectionist models in information retrieval. These models consist of several related information processing approaches, such as neural networks, associative networks, spreading activation models, and parallel distributed processing. Kwok [40] also developed a similar three-layer network of queries, index terms, and documents. A modified Hebbian learning rule will be use to reformulate probabilistic web information retrieval. Rose and Belew [39] extended AIR to a hybrid connectionist and symbolic system called SCALIR which used analogical reasoning to find relevant documents for legal research. Wilkinson and Hingston [41,42] incorporated the vector space model in a neural network for document retrieval. Their network also consisted of three layers: terms, queries, and documents. They have shown that spreading activation through related terms can help improve retrieval performance.

8 CONCLUSION

The World Wide Web (WWW) is a popular and interactive medium will be use to gather, disseminate, and access an increasingly huge amount of information. Researches show that a soft computing technique gives encouraging results and may be effectively used in information retrieval (IR). These reviews prove that it may be possible to use soft computing in infor-

mation retrieval for specific tasks. The literature review above shows soft computing have many attractive applications in information retrieval. However, we believe the application of soft computing can make an information retrieval system more powerful.

REFERENCES

[1] Berners-Lee, Tim, "The World Wide Web: Past, Present and Future", MIT USA, Aug 1996, available at: http://www.w3.org/People/Berners-Lee/1996/ppf.html.

[2] Berners-Lee, Tim, and Cailliau, CN, R., "Worldwide Web: Proposal for a Hypertext Project" CERN October 1990, available at: http://www.w3.org/Proposal.html.

[3] "Internet World Stats. Worldwide internet users", available at: http://www.internetworldstats.com (accessed on Jan 7, 2013).

[4] Bangorn Klabbankoh and Ouen Pinngern, Applied genetic algorithms in information retrieval, IJCIM, 1999

[5] Ahmed A. A. et. al. , Using Genetic Algorithm to Improve Information Retrieval Systems, World Academy of Science, Engineering and Technology, pp. 6-12, 2006

[6] M. Koorangi, K. Zamanifar , A Distributed Agent Based Web Search using a Genetic Algorithm , IJCSNS International Journal of Computer Science and Network Security, 7(1), pp. 65-76, Jan 2007.

[7] Abdelmgeid A. Aly, Applying genetic algorithm in query improvement problem, International Journal "Information Technologies and Knowledge", 7(1), pp. 309-316, 2007

[8] Alin Mihaila , Andreea Mihis, Cristina Mihaila, A Genetic Algorithm for Logical Topic Text Segmentation, pp. 500-505, IEEE 2008

[9] Ziqiang Wang, Xia Sun, Dexian Zhang, Web Document Query Optimization Based on Memetic Algorithm, Pacific-Asia Workshop on Computational Intelligence and Industrial Application., pp. 53-56, IEEE 2008.

[10] Adriano Veloso, Humberto M. Almeida, Marcos Gonçalves, Wagner Meira Jr., Learning to Rank at Query-Time using Association Rules, SIGIR'08, pp. 267-273 , 2008, Singapore.

[11] Rong LI, Jia-heng ZHENG, Chun-qin PEI , Text Information Extraction based on Genetic Algorithm and Hidden Markov Model, First International Workshop on Education Technology and Computer Science, pp. 334-338, 2009.

[12] Habiba Drias, Ilyes Khennak, Anis Boukhedra, Hybrid Genetic Algorithm for large scale Information Retrieval, pp. 842-846, IEEE 2009

[13] Lourdes Araujo and Joaquin Perez-Iglesias, Training a Classifier for the selection of Good Query Expansion Terms with a Genetic Algorithm, IEEE 2010

[14] Pratibha Bajpai and Dr. Manoj Kumar, Genetic Algorithm an Approach to Solve Global Optimization Problems, Indian Journal of Computer Science and Engineering, 1(3), pp. 199-206, 2010

[15] A. S. Siva Sathya and B. Philomina Simon, A Document Retrieval System with Combination Terms Using Genetic Algorithm, International Journal of Computer and Electrical Engineering, 2(1), pp. 1-6, February 2010

[16] A.Leuski, "Evaluating document clustering for interactive information retrieval.," in Proceedings of the 2001 ACM CIKM International Conference on Information and Knowledge Management, Atlanta, Georgia, USA, 2001, pp. 33–44.

[17] M Caramia, G Felici, and A Pezzoli, "Improving search results with data mining in a thematic search engine," Computers & Operations Research, pp. 2387–2404, 2004.

[18] M Gorden, "Probabilistic and genetic algorithms in document retrieval," Communications of the ACM, vol. 31, no. 10, pp. 1208–8, October 1988.

[19] M.Gordon, "User-based document clustering by redescribing subject descriptions with a genetic algorithm," Journal of the American Society for Information Science, vol. 42, no. 5, pp. 311–22, 1991.

[20] F. Dashti and S. A. Zad, "Optimizing the data search results in web using Genetic Algorithm, ―International Journal of Advanced Engineering Sciences and Technologies‖ vol. 1, no. 1, pp. 016 – 022, 2010.

[21] A. Al-Dallal and R.S. Abdul-Wahab, "Genetic Algorithm Based to Improve HTML Document Retrieval," in Developments in eSystems Engineering, Abu Dhabi , 2009, pp. 343 - 348.

[22] Z.S. Ma , and Q.He, "Web Mining: Extracting Knowledge from the World Wide Web," in Data Mining for Business Applications, Longbing Cao et al., Eds.: Springer, 2009, ch. 14, pp. 197-208.

[23] E. Agichtein, E. Brill, and S. Dumais, "Improving web search ranking by incorporating user behavior information," in SIGIR '06 Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, New York, 2006.

[24] Y. Du and H. Li, "An Intelligent Model and Its Implementation of Search Engine," Journal of Convergence Information Technology, vol. 3, no. 2, pp. 57-66, June 2008.

[25] Zacharis Z. Nick and Panayiotopoulos Themis, Web Search Using a Genetic Algorithm, IEEE Internet computing,1089-7801/01 2001, 18-25, IEEE

[26] M.J. Martin-Bautista, H. Larsen, M.A. Vila, A fuzzy genetic algorithm approach to an adaptive information retrieval agent, Journal of the American Society for Information Science 50 (9) (1999) 760–771.

[27] Abe, K.; Taketa, T.; Nunokawa, H. ,"An efficient Information Retrieval Method in WWW using Genetic Algorithms" 1999. Proceedings. 1999 International Workshops on Volume , Issue , 1999 Page(s):522 - 527

[28] P. Pathak, M. Gordon, W. Fan, Effective information retrieval using genetic algorithms based matching functions adaptation, in, Proc. 33rd Hawaii International Conference on Science (HICS), Hawaii, USA, 2000.

[29] Lothar M. Schmitt ," Fundamental Study ,Theory of genetic algorithms", Theoretical Computer Science 259 ,1–61, 2001.

[30] James.F.Frenzel ,"Genetic Algorithms, a new breed of optimization ",IEEE Potentials, 0278-6648/93, IEEE, 1993

[31] Baeza-Yates, R., Ribeiro-Neto, B., Modern Information Retrieval. Addison Wesley, New York, 1999

[32] David E Goldberg ,Genetic Algorithms in Search, Optimization, Machine Learning , Addison Wesley , 1989

[33] Marco Dorigo, Gambardella, Luca Maria, "Ant colonies for the traveling salesman problem." Biosystems, 1997, 43(2): 73-81.

[34] Marco Dorigo, Gambardella, Luca Maria, "Ant colony system: A cooperative learning approch to the traveling salesman problem 1997, 1(1): 53-56

[35] Salton, G. (1989). Automatic text processing. Reading, MA: Addison Wesley.

[36] Maron, M. E., & Kuhns. J. L. (1960). On relevance, probabilistic indexing and information retrieval. Journal qfthe ACM, 7, 2 16-243.

[37] Doszkocs, T. E., Reggia, J., & Lin, X. (1990). Connectionist models and information retrieval. Annual Review of Information Science and Technology (ARIST), 25,209-260.

[38] Belew, R. K. (1989. June). Adaptive information retrieval. In Proceedings of thc twelfth Annual International AiCM/SIGIR Conference on Research and Development in Information Retrieval (pp. 1 I-20). NY, NY: ACM Press

[39] Rose, D. E.. & Belew, R. K. (1991). A connectionist and symbolic hybrid for improving legal research. International Journal ofMan-Machinestudies,35, 1-33.

[40] Kwok, K. L. (1989, June). A neural network for probabilistic information retrieval. In Proceedings of the Twelfth Annual Internationul AC,V/SIGIR Conference on Research and Development in Information Retrieval(pp. 2 I-30). NY, NY: ACM Press.

[41] Wilkinson, R., & Hingston, P. (I 99 1. October). Using the Cosine measure in neural network for document retrieval. In Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in information Retrieval (pp. 202-2 IO).Chicago, IL.

[42] Wilkinson, R.. Hingston. P., & Osborn, T. (1992). Incorporating the vector space model in a neural network used for document retrieval. Library Hi Tech, IO. 69-75.

[43] J. Ilonen, J.-K. Kamarainen, and J. Lampinen. Differential evolution training algorithm for feed-forward neural networks. Neural Processing Letters, 17:93 – 105, 2003.

[44] R. Storn. Differential evolution design of an iir-filter. In IEEE International Conference on Evolutionary Computation, pages 268 – 273, 1996.

[45] R. Storn and K. V. Price. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization, 11(4):314 – 359, December 1997.

[46] Md. Abu Kausar, V S Dhaka and Sanjeev Kumar Singh.: Web Crawler: A Review. International Journal of Computer Applications 63(2):31-36, February 2013. Published by Foundation of Computer Science, New York, USA.