Asian Journal of Information Technology 11 (3): 117-124, 2012
ISSN: 1682-3915

State of the Art Review on Relevance of Genetic Algorithm to Internet Web Search

[1]Kehinde Agbele, [1]Ademola Adesina, [2]Daniel Ekong and [3]Oluwafemi Ayangbekun
[1]Machine Learning and Intelligent Systems Research Group, Department of Computer Science,
University of the Western Cape, Cape Town, South Africa
[2]Department of Mathematical Sciences (Computer Science Option)
Ekiti State University, Ado-Ekiti, Nigeria
[3]College of Information and Communication Technology
Crescent University Abeokuta, Ogun State, Nigeria

Abstract: People use search engines to find information they desire with the aim that their information needs will be met. Information Retrieval (IR) is a field that is concerned primarily with the searching and retrieving of information in the documents and also searching the search engine online databases and internet. Genetic Algorithms (GAs) are robust, efficient and optimization methods in a wide area of search problems motivated by Dawn's principles of natural selection and survival-of-the-fittest. This study presents a novel framework of Information Retrieval Systems (IRS). This study looks at how GAs can be applied in the field of IR and specifically the relevance of Genetic Algorithms to Internet web search. Finally, from the proposals surveyed it turns out that GA is applied to diverse problem fields to internet web search.

Key words: Genetic Algorithm, internet web search, genetic operators, information retrieval, information retrieval systems

INTRODUCTION

There is a virtual explosion in the availability of electronic information. The advent of the Internet or World Wide Web (WWW) has brought far more information than any human being can absorb. The goal of Information Retrieval (IR) systems is to assist user to organize and store such information and retrieve useful information when a user submits a query to the IR systems. To resolve this problem, many research communities have implemented diverse techniques such as full text, inverted index, keyword querying, Boolean querying, knowledge-based, neural network, probabilistic retrieval, genetic algorithm and machine learning. Now, increasing numbers of people use web search engines which enable them to access any kind of information from the Internet in order to formulate better, well-informed decisions. However, the ability of search engines to return useful and relevant documents is not always satisfactory. Often users need to refine the search query several times and search through large document collections to find relevant information. But according to Erba *et al.* (2011) the results returned by the search engine may not be relevant to the users' information need and hence users need to modify and reformulate their queries. The focus of information retrieval is the capability to search for information relevant to individual user's needs within a documents collection which is relevant to the user's query. Baeza-Yates and Ribeiro-Neto (1999) stated that user is in need of information. As discussed by Agbele *et al.* (2010) access to information has important benefit that can be achieved in many areas including social-economic development, education and healthcare. In healthcare for example, access to appropriate information can minimize visits to physicians and period of hospitalization for patients suffering from chronic conditions such as asthma, diabetes, hypertension and HIV/AIDS. Agbele Method examines opening of health information system based on ICT as one fundamental healthcare application area, especially within the context of the Millennium Development Goals to improve the management and quality of healthcare for development at lower cost. It is the responsibility of a user to formulate query and send the query to the search engine (or IRS). Information Retrieval System searches for the matches in the document databases and thus retrieves search results of the matching process. However, based on the relevance, the user will then evaluate and display the

Corresponding Author: Kehinde Agbele, Machine Learning and Intelligent Systems Research Group, Department of Computer Science, University of the Western Cape, Cape Town, South Africa

search results. The relevance of the document is very important to the user. If the user feels that it is a relevant document, he finishes the search else user continues to search in the document database by reformulating the query until the relevant documents are retrieved that will satisfy users' information needs. Genetic Algorithm (GA) (Holland, 1975; Goldberg, 1989) is a probabilistic algorithm simulating the process of natural selection of living organisms and finally come up with an approximate solution to a problem. In GA implementation, the search space is composed of candidate solutions (called individuals or creatures) to an optimization problem evolves a better solutions, each represented by a string is termed chromosome. Each chromosome has an objective function value called fitness. A set of chromosomes together with their associated fitness is called population. This population, at a given iteration of the genetic algorithm is called a generation. In each generation, the fitness of every individual in the population is evaluated from the current population based on their fitness value and modified to form a new population. The new population is then used in the next iteration of the algorithm.

Genetic algorithm terminates when either a maximum number of generations has been produced or a satisfactory fitness level has been reached for the population. If the algorithm has terminated due to a maximum number of generations, a satisfactory solution may or may not have been reached. The working of the genetic algorithm depends upon the constraint how well we choose the initial random keywords.

In the proposed framework for information retrieval as shown in Fig. 1 user gives a mobile SMS-query (Raw query) and the query is reformulated in order to improve the predicted relevance of the retrieved document. The reformulated query is searched against the databases. The proposed retrieval system incorporates the frequency of keyword terms that appear in FAQs databases related to HIV/AIDS content related-documents based on term weighting TFIDF approach by optimizing the ranking of

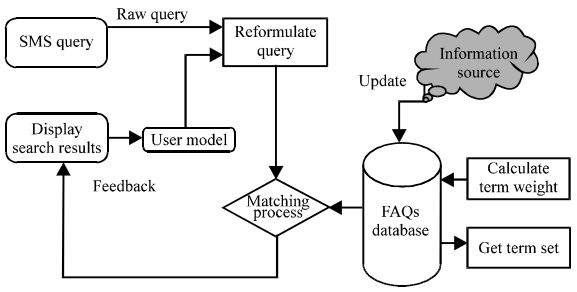

Fig. 1: Information Retrieval System proposed framework

retrieved documents from the search engine. The Information Retrieval System searches for the matches in the document databases and thus retrieves search results of the matching process. Based on the relevance, the user will then display the search results. The relevance of the document is very important to the user. If the user feels that it is a relevant document, he finishes the search else user continues to search in the document database by reformulating the query until the relevant documents are retrieved that will satisfy users' information needs.

Hence, user query reformulations will apply by updating its model. A user model is a stored knowledge about a particular user. Simple model consists usually of keywords describing user's area of interest. Sort those documents according to TFIDF approach. The documents which have the high Retrieval Status Value (RSV) are considered as the top ranked documents.

The two main components in the proposed information retrieval system framework are Document Databases and Reformulated Query Processing System. The document databases stores the databases related to documents and the representations of their information contents based on TFIDF approach. An SMS-query keyword term is also associated with this component which automatically generates a representation for each document by extracting the frequency of the SMS-query keyword terms from the document contents. The reformulated Query Processing System consists of two subsystems: Searching-Matching unit and Displaying-Ranking unit.

Searching unit allows user to search the documents from the document database and matching unit does a comparison of all documents against the user's query. To improve the predicted relevance of the retrieved document, the reformulated query is searched against the databases. Searching-Matching unit does a thorough search and finds out which documents match the user query. This unit retrieves almost all the documents that match either part or whole of the entire query that is the unit retrieves relevant amid non relevant documents. Displaying unit displays the search results based on relevance of the documents to user information needs and ranking unit ranks the document according to the relevance of the user query. Displaying-Ranking unit does a detailed display of search results and find out which documents have high RSV are considered as the top ranked documents. Therefore, Information Retrieval (IR) system ranks the documents according to the RSV between document and the query. If a document has got high RSV that document is closer to the query. In other words the document is relevant to the query.

Generally IR System ranks the list of documents in the descending order. After processing the query effectively, the top most relevant documents are retrieved and it is given to the user. Though, relevance feedback is one of the processes in an information retrieval system that seeks to improve the system's performance based on a user's feedback. It modifies queries using judgments of the relevance of a few, highly-ranked documents and has historically been an important method for increasing the performance of information retrieval systems. Specifically, the user's judgments of the relevance or non-relevance of some of the documents retrieved are used to add new terms to the query and to reweigh query terms. For example if all the documents that the user judges as relevant contain a particular term then that term may be a good one to add to the original query. It is made known (Salton and Buckley, 1990) that relevance feedback has improved the system's overall performance by 60-170% for different document collections. Given the apparent effectiveness of relevance feedback techniques, it is important that any proposed model of information retrieval include these techniques. In the proposed system, rather than modifying the matching function, researchers will modify the query vector using genetic algorithm to adapt the query vectors and to reflect a user's feedback about relevance.

RELEVANCE OF GENETIC ALGORITHM IN INFORMATION RETRIEVAL AND INTERNET WEB SEARCH

In designing GA (Schmitt, 2001), there are three main components have to take into consideration. This research study presents an application of GA as relevant feedback method aiming to adapt keywords weights. In the following, researchers shall give the three main components; the first one is coding the problem solutions, subsequent is to find a fitness function that can optimize the performance and finally, the set of parameters including the population size, population structure and genetic operators. Genetic algorithms are generally used to for solving timetabling (Karova, 2004), stock marketing (Lin *et al*., 2008) and job scheduling (Ying and Bin, 1996) problems.

A Genetic Algorithm (GA) is used as a powerful tool to search solutions in the domain of relevant features and is suitable for the information retrieval for the following justifications (Frenzel, 1993; Boughanem *et al*., 2003). Genetic Algorithms are one of the powerful searching tools known for its robustness and fast search abilities. The document search space represents a high dimensional space. Hence, they are appropriate for information retrieval. The probabilistic discovery induced by Genetic Algorithm permits the discovery of new areas in the document search space are suitable for information retrieval. Therefore, the conventional techniques of relevance feedback are not efficient when no relevant documents are retrieved with the initial query. Genetic Algorithm influences a population of queries rather than a single query, in comparison with the classical information retrieval models. However, each query may retrieve a subset of relevant documents that can be merged. So, Genetic Algorithm is more efficient than using a hill climbing algorithm for effective information retrieval. Genetic Algorithm contributes to maintain useful information links representing a set of keyword terms of the relevant documents. Thus, the conventional techniques of query expansion influence each term independent of others.

A GENERALIZED GENETIC MODEL APPROACH FOR INFORMATION RETRIEVAL

Information Retrieval (IR) is concerned primarily with finding and returning information stored in computers that is relevant to a user's needs (query). With the advent of the Internet and World Wide Web (Web for short), IR has acquired remarkable practical impact as well as theoretical importance. Therefore, IR is an area committed to the management of large collections of information and to the retrieval of helpful information from the documents collection for users (Baeza-Yates and Ribeiro-Neto, 1999) tries to provide the user with easy access to the document databases that will satisfy his information needs and which is relevant in a suitable time interval. IR may be defined, in general as the problem of selection of information from documents collection (storage) in response to search queries issued by a user.

Therefore, the goal of an Information Retrieval System (IRS) is to estimate the relevance of information needs to a user's information expressed in a query. In this regards, IRS process user queries trying to allow the user to access relevant information in a suitable time interval. According to Bookstein (1983) states that basically people want an IRS to provide lists of documents ordered by the importance of documents to them. Relevant documents are to be ranked first than the non-relevant documents so that the user need not search the entire documents collection to look for documents of interest. In general, documents with higher similarity to query are judge more relevant to the query and should be retrieved first. This review includes diverse suggestions found for the relevance of genetic algorithm to the area of information retrieval. Moreover, diverse types of IR problems that are solved by genetic algorithm are analyzed.

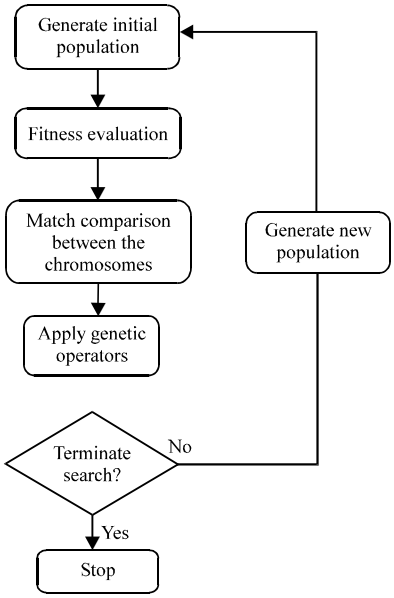

Fig. 2: A Genetic Algorithm approach to information retrieval

The information retrieval approach shown in Fig. 2 can be applied to Genetic Algorithm for retrieving information. In this case, query and documents are represented as gene (chromosome). An initial population of query is generated. The query is sent to the IRS (search engine), a matching process is carried out between the query chromosome and the document chromosome and then the document is considered as relevant. If non-relevant documents are found then, the query is reformulated. The query is reformulated until a relevant document is retrieved. Besides, query will be seen as a vector of a user in the context of the query. The search vectors will be expressed as keyword terms against its associated term weights value. The key primary concern in representation is how to select proper keyword terms. Hence, representation commences by extracting the keyword terms that are considered as content identifiers and classifying them into the arrangement comprised of pairs (k, w) where k is a keyword term and w is the term's weight.

SURVEY OF GENETIC ALGORITHM FOR INTERNET SEARCH

Ever since the advent of the public network Internet, the quantity of available information is rapidly rising. One of the most important uses of this public network is to find information. In such a huge and unstable information collection, today's greatest problem is to find relevant information. It is necessary to improve the existing search agents. Diverse proposals are put forward that use GAs in Internet search with this aim. According to Koorangi and Zamanifar (2007), proposed the problems of existing internet search engines are examined and hence the need for a novel design is warranted. To make search engines work more efficiently, new thoughts on how to improve existing internet engines are presented and then an adaptive technique for internet meta-search engines with a multi-agent especially the mobile agent is presented. In the technique, the understanding between stationary and mobile agents is used as an indication to make it more competence. However, the meta-search engine gives the user needed documents based on the multi-agent mechanism. The combination of the results obtained from the search engines in the network is done in parallel. In this regards, a feedback mechanism gives the meta-search engine the user's suggestions about the found documents which leads to a new query using a genetic algorithm.

Varadarajan *et al.* (2008) proposed a new technique that given a keyword query on the fly generates new pages, called composed pages which include all query keywords. The composed pages are generated by extracting and mending together relevant pieces from hyperlinked Web pages and retaining links to the original Web pages. To rank the composed pages, both the hyperlink structure of the original pages and the associations between the keywords in each page are considered. The proposed technique is used to evaluate heuristic algorithms to efficiently generate top composed pages. Maleki-Dizaji (2003) uses Genetic Algorithms (GAs) for user-modelling of adaptive and exploratory behaviour in an Information Retrieval System. Maleki Dizaji choices of the underlying genetic operators are mentioned as a major drawback in the use of GAs. However, GA is primarily used to solve optimization problems; their use for information retrieval is gaining ground.

Cheng *et al.* (2002) proposed an improved genetic algorithm which solves the issues in two generation competitive genetic algorithm. So, it changes the selection technique of the simple genetic algorithms and improves search efficiency but local best search ability cannot be improved. The proposed algorithm does the adaptive adjustment of the mutation probability and the position of crossover and mutation probability in chromosomes.

Sinha and Chande (2010) surveyed an effective GA that monitor the success of internet database management system by combining functionality, quality and complexity of query optimizer for finding good solutions to the problem. Marghny and Ali proposed a framework for web mining, the applications of data mining

and knowledge discovery techniques to data collected in World Wide Web (WWW) and a genetic search for search engines. The researchers defined an evaluation function that is a mathematical formulation of the user request and to define a steady state Genetic Algorithm (GA) that evolves a population of pages with binary tournament selection. This approach chooses one crossover position within the page randomly and exchanges the link after that position between both individuals (web pages).

Lin *et al.* (2002) proposed a class based internet document management and access system; ACRID, it uses machine learning techniques to organize and retrieve internet documents. The knowledge acquisition process of ACRID automatically learns the classification knowledge from classified internet documents into one or more classes. The two phase search engine in ACRID will use the hierarchical structure for responding to user queries.

Chen *et al.* (2005) proposed and applied a dynamically terminated genetic algorithm to generate page clippings from web search results. The Page Clipping Synthesis (PCS) search method applies a dynamically terminated genetic algorithm to generate a set best of run page clippings in a controlled amount of time. In the proposed approach, the dynamically terminated genetic algorithm yields cost-effective solutions compared with solutions reached by conventional genetic algorithms. Chen *et al.* (1998) proposed an intelligent personal spider approach for Internet searching. The researchers implemented Internet personal spiders based on best first search and genetic algorithm techniques. The used GA applies stochastic selection based on Jaccard's fitness with heuristic-based crossover and mutation operators. These personal spiders dynamically take a set of user's selected starting homepages in the web, based on the existing links and keyword indexing.

Silva *et al.* (2009) proposed the use of Genetic Programming (GP) to derive approach for the combination of three different sources of evidence for ranking documents in the web search engines. The initialization method of the approach defines the method to create the initial population. Two methods can be adopted, grow or full which represent small changes in the algorithm to construct the trees. The approach is useful for coping with search engines that can request diverse forms of queries for submission. A comprehensive comparison of the diverse proposals prepared by different researchers is shown in Table 1.

Table 1: Comparison of Diverse Proposals That Use Genetic Algorithm for Internet Search

Diverse proposals	Reason for GA	Chromosomes	Fitness function adopted	Genetic operators used
Eissa and Alghamdi (2005)	Genetic algorithm is used to optimize the profiles whereas the relevance feedback is used to adapt it	Represent a gene as a term, an individual as a document and the population as the profile	$F(Pi)=\sum_{k} S(D_{Pk}, P_i)/\#D_p$	Selection
Vallim and Coello (2003)	Combines user's feedback to new documents retrieved by the agent with a genetic algorithm	Individuals represented by a query vector and its adaptation rate	$Q = Q + \alpha f$ $F(Q) = f + \beta f$	Two point crossover and mutation operator
Li *et al.* (2000)	Realize the scheduling strategy of agent manager	Search space is represented as weight field in the search engine. Field are search parameters	Adaptation function $\phi(\text{agent}) = \Gamma(f,p,c,u,t)$	One point cross over and single point mutation
Caramia *et al.* (2004)	Select a subset of original pages for which the sum of scores is large	Chromosomes represent subsets of pages of bounded cardinality. Each page is a gene	$Ff(c) = \alpha.t_1(C)+\beta.t_2(C)+\gamma.t_3(C)$	Single point crossover
Marghny and Ali (2005)	Steady state genetic algorithm for optimizing web search	Initial population is generated by heuristic creation operator which queries standard engines to obtain pages	Fitness function evaluates web pages is a mathematical formulation of link quality, page quality and Mean quality function	Binary tournament selection and single point crossover
Chen *et al.* (1998)	GA implemented as a spider to find most relevant home pages in the entire internet	Chromosomes represent all input home pages in a set	Jaccard's coefficient function	Heuristic based cross over and simple mutation
Jin *et al.* (2002)	Improving the searching performance	Initial population represented by binary coding selected at random	$F_d = (F_{max}-F_{min})/F_{aver}$	Fitness proportion selection, adaptive adjusting crossover and mutation operation range
Fan *et al.* (2003)	Genetic programming to the ranking function discovery problem leveraging the structural information of HTML documents	Chromosomes represent html pages	The fitness evaluation of each ranking tree is done at the level multiple queries. $P\_Avg = \sum P_i/TRel$, $P_i = i/Rank_i$	Single point crossover and one point mutation
Koorangi and Zamanifar (2007)	Query reformulation in search engine	Initial population consists of first five keywords of the user dictionary	CHK fitness function	One point crossover and inversion mutation operator

APPLICATIONS OF GENETIC ALGORITHMS IN INFORMATION RETRIEVAL

The information explosion caused by the expansion of the communications has promoted the creation of large information stores. The classification and hierarchization of these data warehouses have become a pressing need, especially in processes of Information Retrieval. Therefore, there has been an increasing interest in the application of genetic algorithm tools to information retrieval in the last few years. The machine learning concept (Mitchell, 1997) whose aim is the design of system able to automatically acquire knowledge by themselves seems to be interesting (Chen *et al*., 1998). Genetic Algorithms are not specifically learning algorithms but also offering a powerful and domain independent search ability that can be used in many learning tasks, since learning and self-organization can be considered as optimization problems in many cases. As a result of this reason, the applications of genetic algorithms to information retrieval have increased in the last few years. Among others in the following researchers shall examine some of the diverse proposals made in these fields in the last few years.

Clustering of document and terms: In this field, two approaches have been applied for obtaining user-oriented document clusters. Robertson and Willet (1994) look for groups of terms appearing with similar frequencies in the documents of collection. The researchers consider a GA grouping the terms without maintaining their initial order. The main features of the GA are given.

Representation scheme: Two different coding schemes are considered to include division-assignment and separator methods.

Initial population: The first generation of the chromosomes depends on the chosen coding and the rest of individual are randomly generated.

Operators: Each operator has an application probability associated and it is selected spinning the roulette. Different crossover and mutation operators are used.

Fitness function: A measure of the relative entropy and Pratt's measure are two proposals adopted.

Matching function learning: The objective of matching function learning is to use a genetic algorithm to generate a similarity measure for a vector space information retrieval system to improve its retrieval efficiency for a defined user. This constitutes new relevance feedback beliefs since matching functions are adapted instead of queries. In this regards, two different variants have been proposed in the specialized literature.

Automatic similarity measure learning: According to Gordon (1988) and Fan *et al*. (2004) introduced a genetic algorithm that automatically learns a matching function with relevance feedback. Besides, the similarity functions are represented as trees and a classical generational scheme, the usual GA crossover are considered.

Linear combination of existing similarity functions: According to Pathak *et al*. (2000) propose a new weighted matching function which is the linear combination of different existing similarity functions. The weighting parameters are estimated by a genetic algorithm based on relevance feedback from users. The researcher use real coding, a classical generational scheme, two-point crossover and Gaussian noise mutation. Finally, the algorithm is tested on the Cranfield collection.

Automatic document indexing: The applications in this area adapt the descriptions of the documents in the documentary base with the aim of facilitating document retrieval in the face of relevant queries. According to Gordon (1988) proposes a Genetic Algorithm to derive the document descriptions. He chooses a binary coding scheme where each description is a fixed length and a binary vector. The genetic population is composed of diverse descriptions for the same document. The fitness function is based on calculating the similarity between the current document description and each of the queries (for which the document is relevant or non-relevant) by means of the Jaccard's index and then computing the average adaptation values of the description for the set of relevant and non-relevant queries.

In Gordon research, GA considered is quite unusual as there is no mutation operator and the crossover probability is equal to 1. With regard to the selection scheme, the number of copies of each chromosome in the new population is calculated and dividing its adaptation value by the population average. Also Fan *et al*. (2000) propose an algorithm for indexing function learning based on GA whose aim to obtain an indexing function for the key term weighting of a documentary collection to improve the information retrieval process.

Query learning: This is the most extended group of applications of genetic algorithms in information retrieval. Every proposal in this group use genetic algorithms either like a relevance feedback method or like an Inductive

Query By Example (IQBE) algorithm. The fundamental of relevance feedback lies in the fact that either user normally formulates queries composed of terms which do not match the terms (which used to index the relevant documents to their needs) or they do not provide the appropriate weights for the query terms. The operation mode is involving and modifying the previous query (adding and removing terms or changing the weights of the existing query terms) with taking into account the relevance judgements of the documents retrieved by it, constitutes a good way to solve the latter two problems and to improve the precision and especially the recall of the previous query (Van Rijsbergan, 1979).

Therefore, IQBE was proposed as a process in which searchers provide sample documents (examples) and the algorithms induce (or learn) the key concepts in order to find other relevant documents (Chen *et al*., 1998). This technique is a process for assisting the users in the query formulation process performed by machine learning techniques. It works by taking a set of relevant (and optionally, non-relevant documents) provided by a user and applying an off-line learning process to automatically generate a query describing the user's information needs. Besides, Smith and Smith (1997) propose a Genetic Algorithm for learning queries for Boolean Information Retrieval System. Although, the researchers introduce concept approach as a relevance feedback algorithm, the experimentation is actually closer to IQBE framework.

According to Yang and Korfhage (1994) propose a similar Genetic Algorithm to that of Robertson and Willet (1994). They use a real coding with the two-point crossover and random mutation operators (besides, crossover and mutation probabilities are changed throughout the GA run). The selection is based on a classic generational scheme where the chromosomes with a fitness value below the average of the population are eliminated and the reproduction is performed by Baker's mechanism.

CONCLUSION

This review has dealt with the fundamentals of the information retrieval and genetic algorithm. Issues that can be solved using Genetic Algorithm and research areas in internet web search are discussed in this study. It also deals with diverse proposals in internet web search which are promising and growing research areas. This study examines the relevance of genetic algorithm in diverse fields of internet web search, some applications of genetic algorithms to information retrieval and a survey of the research works done in internet web search area has been examined carefully and the results so far have thus been very promising and encouraging. Though, this research project is at development and implementation stage. It is the strong belief that the full implementation and evaluation of the proposed information retrieval systems will assist users in documents ranking according to their relevance. Therefore, HIV/AIDS content-related documents with higher similarity query are to be judged more relevant to the SMS-query keyword terms and should be retrieved first using genetic algorithm to adapt the documents weight via relevance feedback from users. This will in turn help HIV/AIDS managements and lower the cost of healthcare provision.

REFERENCES

Agbele, K., H. Nyongesa and A. Adesina, 2010. ICT and information security perspectives in E-health systems. J. Mobile Commun., 4: 17-22.

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

Bookstein, A., 1983. Outline of a general probabilistic retrieval model. J. Document., 39: 63-72.

Boughanem, M., C. Chrisment and L. Tamine, 2003. Multiple query evaluation based on an enhanced genetic algorithm. Inform. Process. Manage., 39: 215-231.

Caramia, M., G. Felici and A. Pezzoli, 2004. Improving search results with data mining in a thematic search engine. Comput. Operat. Res., 31: 2387-2404.

Chen, H., C. Yi-Ming, M. Ramsey and C. Yang, 1998. An intelligent personal spider (agent) for dynamic Internet/Intranet searching. Decision Support Syst., 23: 41-58.

Chen, L.C., C.J. Luh and C. Jou, 2005. Generating page clippings from web search results using a dynamically terminated genetic algorithm. Inform. Syst., 30: 299-316.

Cheng, J., W. Chen, L. Chen and Y. Ma, 2002. The improvement of genetic algorithm searching performance. Proce. Int. Conf. Mach. Learn. Cybern., 2: 947-951.

Erba, F.G., Z. Yu and L. Ting, 2011. Using explicit measures to quantify the potential for personalizing search. Res. J. Inform. Technol., 3: 24-34.

Fan, W., M. Gordon and P. Pathak, 2004. Discovery of context-specific ranking functions for effective information retrieval using genetic programming. IEEE Trans. Knowledge Data Eng., 16: 523-527.

Fan, W., M.D. Gordon and P. Pathak, 2000. Personalization of search engine services for effective retrieval and knowledge management. Proceedings 2000 International Conference on Information Systems (ICIS'2000), Brisbane, Australia, pp: 20-34.

Fan, W., M.D. Gordon, P. Pathak, W. Xi and E.A. Fox, 2003. Ranking function optimization for efficient web search by genetic programming. An Empirical Study Department of Computer Science of Virginal Tech, Michigan, Florida Universities.

Frenzel, J.F., 1993. Genetic algorithms. IEEE Potentials, 7: 21-24.

Goldberg, 1989. Genetic Algorithms. Pearson Education India, India, ISBN: 9788177588293.

Gordon, M., 1988. Probabilistic and genetic algorithms in document retrieval. Commun. ACM, 31: 1208-1218.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. 1st Edn., University of Michigan Press, Ann Arbor, Michigan, ISBN: 0472084607.

Karova, M., 2004. Solving timetabling problems using genetic algorithms, in: Electronics technology: Meeting the challenges of electronics technology progress. Int. Spring Seminar, 1: 96-98.

Koorangi, M. and K. Zamanifar, 2007. A distributed agent based web search using a genetic algorithm. Int. J. Comput. Sci. Network Security, 7: 65-76.

Li, W., B. Xu, H. Yang, W. Cheng-Chung and C.W. Lu, 2000. Application of genetic algorithm in search engine. Proceedings of the International Conference on Microelectronic Systems Education, November 13-13, 2000, USA., pp: 366-371.

Lin, L., L. Cao, J. Wang and C. Zhang, 2008. The applications of genetic algorithms in stock market data mining optimization. Capital Market CRC, Sydney 2000, Australia.http://www.silicon.com/white-papers/view/data-tools/fulltext-60443133/.

Lin, S.H., J.M. Ho and Y.M. Huang, 2002. ACRID, intelligent internet document organization and retrieval. Trans. Knowl. Data Eng., 14: 559-613.

Maleki-Dizaji, S., 2003. Evolutionary learning multi-agent based information retrieval systems. Ph.D. Thesis, Sheffield Hallam University.

Mitchell, T., 1997. Machine Learning. McGraw-Hill, New York.

Pathak, P., M. Gordon and W. Fan, 2000. Effective information retrieval using genetic algorithms based matching functions adaptation. Proceedings of the 33rd Hawaii International Conference on System Sciences, January 4-7, 2000, Hawaii, USA.

Robertson, A.M. and P. Willet, 1994. Generation of equifrequent groups of words using a genetic algorithm. J. Document., 50: 213-232.

Salton, G. and C. Buckley, 1990. Improving retrieval performance by relevance feedback. J. Am. Soc. Inform. Sci., 41: 288-297.

Schmitt, L.M., 2001. Fundamental study, theory of genetic algorithms. Theor. Comput. Sci., 259: 1-61.

Silva, T.P.C., E.S. de Moura, J.M.B. Cavalcanti, A.S. da Silva, M.G. de Carvalho and M.A. Goncalves, 2009. An evolutionary approach for combining different sources of evidence in search engines. Inform. Syst., 34: 276-289.

Sinha, M. and S.V. Chande, 2010. Query optimization using genetic algorithms. Res. J. Inform. Technol., 2: 139-144.

Smith, M.P. and M. Smith, 1997. The use of genetic programming to build Boolean queries for text retrieval through relevance feedback. J. Inform. Sci., 23: 423-431.

Vallim, M.S. and J.M.A. Coello, 2003. An agent for web information dissemination based on a genetic algorithm, systems, man and cybernetics. Int. Conf., 4: 3834-3839.

Van Rijsbergan, C., 1979. Information Retrieval. 2nd Edn., Buttersmiths, London, Pages: 224.

Varadarajan, R., V. Hristidis and T. Li, 2008. Beyond single-page web search results. Trans. Knowledge Data Eng., 20: 411-424.

Yang, J. and R. Korfhage, 1994. Query modifications using genetic algorithms in vector space models. Int. J. Expert Syst., 7: 165-191.

Ying, W. and L. Bin, 1996. Job-shop scheduling using genetic algorithm. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, October 14-17, 1996, Beijing, China, pp: 1994-1999.