International Journal of Computer Theory and Engineering, Vol. 1, No. 4, October2009
1793-8201

Review on Applicability of Genetic Algorithm to Web Search

S.Siva Sathya and Philomina Simon, *Member IACSIT*

***Abstract*—Information Retrieval (IR) is concerned with searching and retrieving information within the documents and also searching the online databases and internet. Genetic Algorithms (GA) are robust and efficient search and optimization techniques inspired by the Darwin's theory of natural evolution. In this paper, a novel frame work of information retrieval system is proposed. The applicability of Genetic algorithms in the field of web search and a review on how a GA is applied to different problem domains in web search is discussed.**

***Index Terms*—Genetic algorithm, Information retrieval, Genetic Operators, Web Search, Information retrieval system**

I. INTRODUCTION

The goal of an Information Retrieval System (IRS) is to help a user to locate the most similar documents that have the potential to satisfy the user information needs. To solve this problem, researchers have implemented several methods such as inverted index, Boolean querying, knowledge-based, neural network, probabilistic retrieval, genetic algorithm and machine learning approach.

The focus of information retrieval is the ability to search for information relevant to a user's needs within a collection of data which is relevant to the user's query.[31] User is in need of information. User will formulate query and send the query to the information retrieval system. Information retrieval system searches for the matches in the document database and retrieves results. The user will then evaluate the results based on the relevance. If the user feels that it is a relevant document, he finishes the search else user continues to search in the database by reformulating the query until the relevant documents are retrieved.

Genetic algorithm [34] is a simulation technique that uses a formal approach to simulate above situation and finally come up with an approximate solution to a problem. Genetic algorithms are implemented as a computer simulation in which a population of abstract representations (called chromosomes or the genotype or the genome) of candidate solutions (called individuals, creatures, or phenotypes) to an optimization problem evolves toward better solutions. The solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible. The evolution usually starts from a population of randomly generated individuals and happens in generations. In each generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness), 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 summary of the web search is as follows. Start with some random keywords. Search for those keywords. Select "acceptable" results from your search results and mark down some keywords from it. Do this until the results are approximately what you were looking for. Stop if you are searching over and over for many times but you are not getting good results. The working of the genetic algorithm depends upon the constraint how well we chose our initial random keywords.

Section 2 discusses about the relevance of genetic algorithm in the field of Web search. Section 3 covers about the novel framework for information retrieval. Section 3 deals with the review of different research works and proposals in the area of web search where GA is used.

II. A INFORMATION RETRIEVAL SYSTEM – A PROPOSED FRAMEWORK

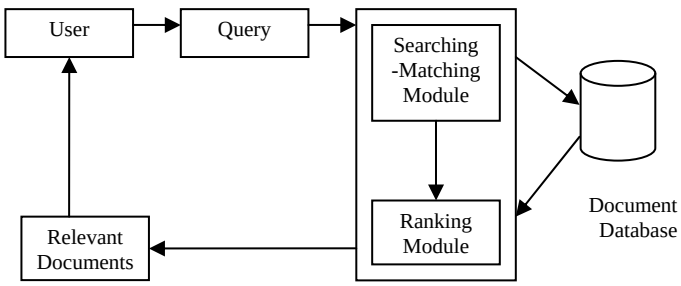

Fig 1. Information retrieval system – proposed framework

In the proposed frame work for information retrieval shown in Fig 1, user gives a query and the query is searched against the document database. The query is compared against the documents and a similarity measure is calculated to find out whether that particular document is relevant to the query or not. If the document is relevant, it is stored in a buffer. After retrieving the relevant documents from the database, Sort those documents according to the similarity measure. The documents which have the high similarity measure value are considered as the top ranked documents.

F. A. S. Siva Sathya is with the Pondicherry University, Department of Computer Science, India

S. B. Philomina Simon is with the Pondicherry University, Department of Computer Science, India

The two main components in the proposed IRS frame work are Document Database and Query Processing System. Document Database stores the documents and the representations of their information contents. An indexer is also associated with this component which automatically generates a representation for each document by extracting the document contents. The query processing System consists of two subsystems: Searching-Matching Module and Ranking Module.

Searching module allows user to search the documents from the document database and matching module does a comparison of all documents against the user's query. Searching-Matching module does an exhaustive search and finds out which all documents match the user query. This module retrieves almost all the documents that match either part or whole of the full query, i.e. the module retrieves relevant as well as non relevant documents. Ranking module ranks the document according to the relevance of the user query. IR system ranks the documents according to the similarities between document and the query. If a document has got high value of similarity, that document is closer to the query. In other words the document is relevant to the query.

Usually 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.

A. Relevance of Genetic Algorithm in Information Retrieval and Web Search

There are three main components that have to be taken care while designing GA[22]. The first one is coding the problem solution, next 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. GAs are widely used for solving job scheduling [32] and time tabling [33] problems.

Genetic algorithms is a powerful search mechanism and it is suitable for the information retrieval for the following reasons [27], [23]

- The document search space represents a high dimensional space. GAs are one of the powerful searching mechanism known for its robustness and quick search capabilities. So they are suitable for information retrieval.
- In comparison with the classical information retrieval models, GA manipulates a population of queries rather than a single query. Each query may retrieve a subset of relevant documents that can be merged. GA is more efficient than using a hill climbing algorithm.
- The traditional methods of query expansion manipulate each term independent of other. GA contributes to maintain useful information links representing a set of terms indexing the relevant documents.
- The traditional methods of relevance feed back are not efficient when no relevant documents are retrieved with the initial query. The probabilistic exploration induced by GA permits the exploration of new areas in the document space.

B. Genetic Model For Information Retrieval – a Generalized Approach

Information retrieval (IR) tries to make a suitable use of document databases, allowing the users access to the information which is really relevant in an appropriate time interval [31]. Information Retrieval (IR) may be defined, in general, as the problem of selection of documentary information from storage in response to search questions provided by a user. Information retrieval system (IRS) deals with documentary bases containing textual, pictorial or vocal information and process user queries trying to allow the user to access relevant information in an appropriate time interval. This survey includes different proposals found for the application of genetic algorithm to the field of information retrieval. Different kinds of IR problems that are solved by genetic algorithms are analyzed.

This information retrieval approach given in Fig 2 can be applied to GA for retrieving information. Query and documents are represented as chromosome. An initial population of query is created. The query is sent to the information retrieval system, a match is done 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. The document collection consists of many documents containing information about various subject or topics of interests. The primary concern in representation is how to select proper index terms. Representation proceeds by extracting the key words that are considered as content identifiers and organizing them into the format.

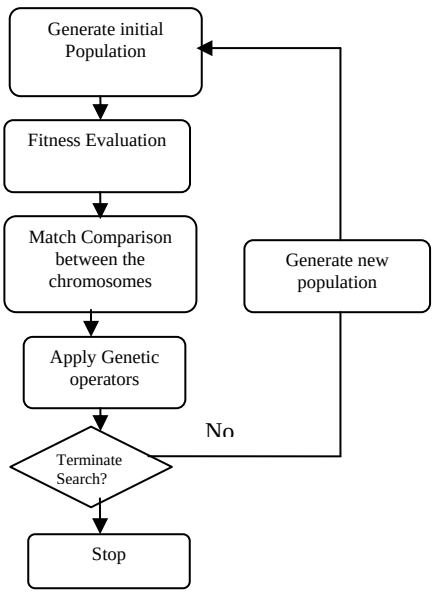

Fig.2. A genetic algorithm approach to information retrieval.

III. Genetic Algorithm for Web Search

The Internet services are popular and appealing to many on-line users, difficulties with search are expected to worsen as the amount of on-line information increases. It is necessary to improve the existing search agents. Different proposals are put forward by the authors that use GAs in Internet search with this aim.

In [7], an intelligent personal spider approach for Internet searching is proposed. Chen et al. implemented Internet personal spiders based on best first search and GA 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 and search for the most closely related homepages in the web, based on the existing links and keyword indexing.

Shian-Hua Lin, Jan-Ming Ho, Yueh-Ming Huang [24] proposed a class based internet document management and access system : ACRID, it uses machine learning techniques to organize and retrieve internet documents. 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.

Pablo Castells, Miriam Ferna´ndez, and David Vallet [25] proposed a model which will retrieve the information on the semantic web. He exploited ontology based knowledge bases to improve the search over large document repositories. Semantic search is combined with traditional key word based retrieval to achieve tolerance to knowledge base incompleteness.

Qiao lin, Huang wei-tong, Wen qi, Fu xiao.long [26] proposed an integrated information retrieval support system (IIRSS) for multiple distributed heterogeneous cross-lingual information sources. He introduced the three-layer framework of the IIRSS and discussed a multiagent-based information collection technique. Proposed integrated information retrieval support system, the IIRSS can help Web search engines retrieve cross-lingual information from heterogeneous resources stored in multi-databases in a campus network.

Ramakrishna Varadarajan, Vagelis Hristidis, and Tao Li [28] proposed a new method that, given a keyword query, on the fly generates new pages, called composed pages, which contain all query keywords. The composed pages are generated by extracting and stitching 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 within each page are considered. The proposed method is used to evaluate heuristic algorithms to efficiently generate top composed pages.

Zacharis Z. Nick and Panayiotopoulos Themis[1] proposed Webnaut , an intelligent agent system which uses a genetic algorithm, which enables it to learn the user's interests and to adapt as user interests change over time. The learning process is driven by user feedback to the agent's filtered selections. The GA agent collects and evaluates new HTML pages from the Web, using information included in examples provided by the user. Pages that score high are served to the user by the learning agent. Judit Bar-Ilan studied to measure how similar are the rankings of search engines on the overlapping results and compared compare rankings of results for identical queries retrieved from several search engines [29]. The method is based only on the set of URLs that appear in the answer sets of the engines being compared. For comparing the similarity of rankings of two search engines, the Spearman correlation coefficient is computed. The findings after the study indicate that the large public search engines on the Web employ considerably different ranking algorithms.

Lin-Chih Chen, Cheng-Jye Luh, Chichang Jou proposed applied a dynamically terminated genetic algorithm to generate page clippings from web search results. The page clipping synthesis [18] (PCS) search method applies a dynamically terminated genetic algorithm to generate a set of 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.

Jin Cheng, Wei Chen ,Li Chen, Yao Ma proposed an improved genetic algorithm [13] which solves the issues in two generation competitive genetic algorithm .In two generation algorithm it changes the selection method of the simple genetic algorithms and improves search efficiency, but local best search ability can't be improved. The proposed algorithm does the adaptive adjustment of the mutation probability and the position of cross over and mutation probability in chromosomes.

M. H. Marghny and A. F. 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 [10] for search engines. They 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. Querying standard search engine performs the creation of individuals. The crossover operator with probability of crossover Pc is performed by selecting two parent individuals (web pages) from the population. It chooses one crossover position within the page randomly and exchanges the links after that position between both individuals (web pages).

M. Koorangi, K. Zamanifar[12] proposed the problems of current web search engines are analyzed, and the need for a new design is justified. Innovative ideas on how to improve current web search engines are presented, and then an adaptive method for web meta-search engines with a multi-agent specially the mobile agents is presented to make search engines work more efficiently. In the method, the cooperation between stationary and mobile agents is used to make more efficiency. The meta-search engine gives the user needed documents based on the multi-stage mechanism. The merge of the results obtained from the search engines in the network is done in parallel. 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

Catalogues play an important role in most of the current web search engines. In [30], Loia and Luengo present 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 (clustering fitness and quality of distribution). As genetic operators, the one-point crossover and five different mutation operators (Cutting, Merging, Specialization Grade, Exchange Parent and Change Parent) are defined.

A detailed comparison of the different proposals made by different authors are summarized in Table 1.

TABLE I: COMPARISON OF DIFFERENT PROPOSALS THAT USE GENETIC ALGORITHM FOR WEB SEARCH

Different Proposals	Purpose of GA	Chromosomes	Fitness function	Genetic Operators
Zacharis .Z.Nick Panayiotopoulos Themis[1]	To collect and evaluates new HTML pages from the web, using information included in examples provided by the user.	Vector of random unique keywords from the dictionary ,is created during the initialization of the population	$sim(D_j,D_s)=\frac{\sum_{k=1}^{N} W_{jk}W_{sk}}{\sqrt{\sum_{k=1}^{N} W_{jk}^2 \sum_{k=1}^{N} W_{sk}^2}}$	One point cross over Inversion mutation operator
Maria J, Henrik.L. Larsen et.al [2]	Adaptive internet information retrieval	Each gene represents a fuzzy subset of the document set by means of a Keyword (term) and number of occurrences in a document.	$f_i^j = f_i^{j-1} + P_i^j - L_i^j$ where P^i j and L^i j is the pay off of life tax and chromosome number respectively	Random selection Double point Crossover Random Mutation
Toshihiro Taketa, Horoshi [3]	For evolving information retrieval agents	Genes are represented by the search parameters	$F=\left(\frac{SH}{MH}+\frac{SI}{MI}\right)X\left(1-\frac{ST}{TL}\right)+\left(1-\frac{ME}{MM}\right)$	Selection uses ranking strategy Uniform cross over Single point mutation
Weiguo Fan,Gordon, Pathak [4]	GP approach in problem of automatic term weighing	Initial population consists of terminals and functions. Ramped half and Half method used	Assign a tree retrieval status value by tree representing the current individual	Reproduction Single point Crossover
M.Caramia,G.Felici , A.Pezzoli[5]	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
Arben Asllani, Alireza Lari[6]	Model driven decision support systems for website optimizations	Each generation consists of a vector of sequences of POPSIZE—the population size.	$F=\left(\frac{\sum_{k=1}^{m} D(k)}{M\sum_{k=1}^{m} D(k)*m}\right)*w_1+\left(1-\frac{\sum_{k=1}^{m} V(k)}{M\sum_{k=1}^{m} V(k)*m}\right)*w_2+\left(1-\frac{\sum_{k=1}^{m} P(k)}{M\sum_{k=1}^{m} P(k)*m}\right)*w_3$	Random Cross over Swap Mutation
Chen,Chung,Marshall, Christopher [7]	GA implemented as a spider to find most relevant home pages in the entire internet.	Chromosomes represents all input home pages in a set.	Jaccard's coefficient function	Heuristic based cross over Simple mutation
Weifeng, Hongi, William, Chih[8]	Realize the scheduling strategy of agent manager.	Search space is represented as weight field in the search engine Field are search parameters	Adaptation function $\Phi(agent)=\Gamma(f,p,c,u,t)$	One point Cross over Single point Mutation
Rocio, Carlos, Ana, Nelida [9]	Evolving high quality queries	Chromosome is represented as a list of terms where each term corresponds to a gene	$Fitness(q)=\max(\sigma(c,d_i))$ $d_i \epsilon A_q$	Roulette Wheel Selection Single point cross over One Point mutation
M.H Marghny, A.F Ali[10]	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 Single point cross over
F.Picarougne, Monmarce, Oliver[11]	Genetic search strategy GeniMiner	Chromosome is represented as a web page	Fitness function represents user query and requires to download the page and analyze its content	Creation Operator Singe point mutation

Koorangi, Zamanifar [12]	Query reformulation in a search engine	Initial population consists of first five keywords of the user dictionary	$sim(D_D, D_S) = \frac{\sum_{k=1}^{N} W_{Dk} W_{Sk}}{\sqrt{\sum_{k=1}^{N} W_{Dk}^2 \sum_{k=1}^{N} W_{Sk}^2}}$	One point cross over Inversion mutation operator
Jin,Wei, Li, Yao [13]	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 cross over Mutation operation range
Thomaz Edleno , Joa˜o et.al [14]	GP to derive approach for the combination of different sources of evidence for ranking documents in Web search engines.	Initialization method: 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.	$b_{preq} - 10 = \frac{1}{R}\sum_{r=1}^{R} 1 - \frac{Irrevelant_g(r)}{R+10}$ bpref for informational queries $MRR(QS) = \frac{\sum_{vq_i \in QS} \frac{1}{PosRelAns(q_i)}}{\lvert QS \rvert}$ Mean Reciprocal Ranking for navigational queries	greedy-over-selection method. Random cross over
Roc´ıo ,Carlos, Ana et.al [15]	optimization techniques based on Genetic Algorithms to evolve “good query terms” in the context of a given topic.	initial population of queries come from the initial topic description	Fitness(q)=max(σ(c,d_i)) diϵA_q	Roulette Wheel Selection Single point cross over Mutation Pool
Veljko Dragana Jelena Mirkovic[16]	Genetic search algorithms enable intelligent and efficient Internet searches.	Chromosomes represent set of input Web sites given by a user.	Jaccard’s Function	Topic mutation, Spatial mutation, Temporal mutation,
Fan, Gordon, Pathak, Fox[17]	Genetic Programming (GP) to the ranking function discovery problem leveraging the structural information of HTML documents.	Chromosome represents html pages	The fitness evaluation of each ranking tree is done at the level of multiple queries. $p\_avg = \sum P_i / T_{Rel}, P_i = i / Rank_i$	Single point Cross over One point Mutation
Lin-Chih, Cheng, Chichang [18]	Genetic algorithm formulation of page clipping generation	Population size represents the number of all page clippings produced in that generation for a query. chromosomes represent one page clipping.	$f_{pc,ts} = \frac{\sum_{\forall p \text{ in } pc}(G_{p,pc} \times W_{p,pc,ts})}{\log T_{pc}}$,	Roulette Wheel Selection Two point Cross Over Two point Mutation
F. Eissa, H. Alghamdi [19]	Genetic algorithm is used to optimize the profile 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.	$F(Pi) = \frac{\sum_k S(D_{Pk}, P_i)}{\#D_p}$	Selection
Max Streicher Vallim Juan Manuel Adán Coello [20]	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:	$Q = Q + \alpha f$ $f(Q) = f(Q) + \beta f,$	Two point Cross over Mutation Operator

André L. Vizine1, Leandro N. de Castro1 & Ricardo R. Gudwin[21]	The search agent uses a GA to optimize the search of new papers for a group of users instead of a single user. In the GA used here for web search and information filtering, there are two populations of chromosomes to be coevolved.	For first population, chromosome is a pre-defined number of words randomly chosen from the keywords library of each folder. In second population contains the same number of genes of the first population and same number of individuals of the previous generation.	$sim(D_D, D_Q) = \frac{\sum_{k=1}^{N} W_{Dk} W_{Qk}}{\sqrt{\sum_{k=1}^{N} W_{Dk}^2 \sum_{k=1}^{N} W_{Qk}^2}}$,	Binary tournament selection Single point cross over

IV. Conclusion

This survey has dealt with the basics of the information retrieval and genetic algorithm The research areas in web search and various issues that can be solved using GA is discussed in this paper. It also deals with the different proposals in web search which are emerging research areas. This study discusses the applicability of genetic algorithm in different areas of web search and a review of the research works done in web search domain has been discussed.

References

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

[2] 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.

[3] 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

[4] 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.

[5] M. Caramia, G. Felici, A. Pezzoli , Improving search results with data mining in a thematic search engine , Computers & Operations Research 31 (2004) ,2387–2404

[6] Arbe Asllani, Alireza Lari, Using genetic algorithm for dynamic and multiple criteria website optimization ,European Journal of Operational Search 176(2007)1767-1777

[7] H. Chen, C. Yi-Ming, M. Ramsey, C. Yang, An intelligent personal spider (agent) for dynamic Internet/Intranet searching, Decision Support Systems 23 (1998) 41–58.

[8] Weifeng, Yang, Chung, Wei Lu, "Application of genetic algorithm in search engine",IEEE,366-371, 2000

[9] Rocı´o L. Cecchini , Carlos M. Lorenzetti , Ana G. Maguitman, Ne´lida Beatrı´z Brignole , Using genetic algorithms to evolve a population of topical queries , Information Processing and Management (2008) doi,10.1016/j.ipm.2007.12.012

[10] M. H. Marghny and A. F. Ali Web mining based on genetic algorithm, AIML 05 Conference, 19-21 December 2005, CICC, Cairo, Egypt

[11] F.Picarougne,N,Monmarche," Web mining with a genetic algorithm ", International Conference Proceedings, France ,2002

[12] 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), 2007

[13] Jin Cheng, Wei Chen ,Li Chen, Yao Ma , The improvement of genetic algorithm searching performance ,Proceedings of the first International Conference on Machine Learning and Cybernetics , Beijing , 0-7803-7508-4/02 ,947 – 951 ,2002

[14] Thomaz , EdlenoSilvadeMoura ,Joao AltigranS. daSilva et.al ,"An evolutionary approach for combining different sources of evidence in search engines, Information Systems , Vol 34(2) ,April 2009 ,276-289

[15] Roc´ıo L. Cecchini†, Carlos M. Lorenzetti‡ et.al ,"Genetic Algorithms for Topical Web Search: A Study of Different Mutation Rates"

[16] Milutinovic, V. Cvetkovic, D. Mirkovic, J. "Genetic Search Based on Multiple Mutations"Computer, IEEE, 118-119,Nov 2000

[17] W. Fan, M. D. Gordon, P.Pathak, W. Xi, E. A. Fox, "Ranking Function Optimization for Efficient Web Search By Genetic Programming , An Empirical Study", Dept. of Computer Science of Virginia Tech, Michigan, Florida Universities, 2003.

[18] Lin-Chih Chen, Cheng-Jye Luh, Chichang Jou , Generating page clippings from web search results using a dynamically terminated genetic algorithm, Information Systems 30 (2005) 299–316

[19] F.Eissa, H. Alghamdi," Agent Based Information Retrieval System" ,International Conference Proceedings, 265-279,2005

[20] Vallim, M.S.; Coello, J.M.A.," An Agent for Web Information Dissemination Based on a Genetic Algorithm" Systems, Man and Cybernetics, 2003. IEEE ,International Conference on Vol 4, 5-8, 3834-3836,2003.

[21] Vizine, A.L.; de Castro, L.N.; Gudwin, R.R "An Evolutionary Algorithm to Optimize Web Document Retrieval ",International Conference on Integration of Knowledge Intensive Multi-Agent Systems, 273 – 278, 2005.

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

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

[24] Shian-Hua Lin, Jan-Ming Ho, Yueh-Ming Huang ,ACRID ,intelligent internet document organization and retrieval ,IEEE Transactions on Knowledge and data engineering, 14(3),559-613, 2002

[25] D.H. Kraft, F.E. Petry, B.P. Buckes, T. Sadasivan, Genetic algorithm for query optimization in information retrieval: relevance feedback, in: E. Sanchez, T. Shibata, L.A. Zadeh (Eds.), Genetic Algorithms and Fuzzy Logic Systems, 1997, pp. 155–173.

[26] E. Sanchez, H. Miyano, J. Brachet, Optimization of fuzzy queries with genetic algorithms. Applications to a database of patents in biomedical engineering, in: Proc. VI IFSA Congress, Sao- Paulo, Brazil, 1995, pp. 293–296.

[27] M. Boughanem, C. Chrisment, L. Tamine, "Multiple query evaluation based on an enhanced genetic algorithm", Information Processing and Management 39 , 215–231, 2003

[28] Ramakrishna Varadarajan, Vagelis Hristidis, and Tao Li , Beyond Single-Page Web Search Results,IEEE Transactions on knowledge and data engineering, 20(3) ,411 - 424, 2008

[29] Judit Bar-Ilan, Comparing rankings of search results on the Web, Information Processing and Management 41 (2005) 1511–1519

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

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

[32] Wu Ying and Li Bin , Job- shop scheduling using genetic algorithm, IEEE,1994-1999, 1996

[33] Milena karova , Solving timetabling problems using genetic algorithms, IEEE, 96- 98 2004

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