A Detailed Study on Information Retrieval using Genetic Algorithm

Md. Abu Kausar
Department of Computer & System Sciences
Jaipur National University, Jaipur, India
kausar4u@gmail.com

Md. Nasar
School of Computing Science & Engineering
Galgotias University, Gr. Noida, India
nasar31786@gmail.com

Sanjeev Kumar Singh
Department of Mathematics
Galgotias University, Gr. Noida, India
sksingh8@gmail.com

***Abstract*— Information Retrieval (IR) deals with searching, retrieving and presenting information within the WWW and online databases and also searches the web documents. Genetic Algorithms (GA) are robust and efficient search and optimization techniques inspired by the Darwin's theory of natural evolution. In this paper, the applicability of Genetic algorithm (GA) in the field of information retrieval and a review on how a Genetic Algorithm is applied to different problem domains in information retrieval is discussed.**

***Index Terms*— genetic algorithm, information retrieval, genetic operators, Web Search, information retrieval system**

I. Introduction

Information retrieval dealt with the representation, storage, organization, and access to information items. The representation and association of the information items will provide the user with effortless access to the information in which he will be interested. Unfortunately, characterization of the client information need is not a simple problem. Find all the pages containing information on college cricket teams which are maintained by a university in the India and participate in the IPL tournament. To be relevant, the page must include information on the national ranking of the team in the last four years and the email or mobile number of the team coach. Clearly, this complete description of the player information need not be used directly to request information using the recent Web search engines. The user must first translate this information need into a query which will be processed by the IR system. In its most general form, this translation gives a set of keywords which summarize the description of the user information needed. Given the user query, the key goal of IR system is to retrieved information which may be relevant to the user.

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 [1] (see Fig. 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. Fig. 1: illustrates Internet Users in the World Regions.

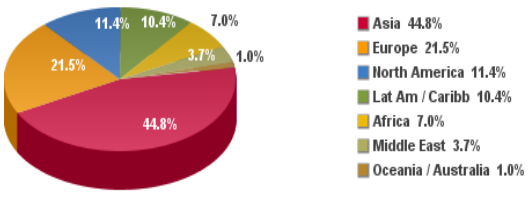

Figure 1. Internet Users in the World Regions (Source: http://www.internetworldstats.com accessed on Jan 7, 2013)

A. Information Versus Data Retrieval

Data retrieval, in the context of an IR system, consists mostly of determining which documents of a collection include the keywords in the user query which, most often, is not sufficient to satisfy the user information need. Actually, the user of an IR system is concerned more with retrieving information about a topic than with retrieving data which satisfies a given query. A data retrieval language aims at retrieving all objects which satisfy

Manuscript received June 1, 2013; revised August 1, 2013.

doi: 10.12720/jiii.1.3.122-127

clearly defined conditions such as those in a regular expression or in a relational algebra expression. Thus, for a data retrieval system, a single wrong object among a thousand retrieved objects means total failure. For an IR system, however, the retrieved objects might be wrong and small errors are likely to go ignored. The major reason for this difference is that IR generally deals with natural language text which one is not always well structured and could be semantically unclear. On the other way, a data retrieval system such as a relational database deals with data that have a well defined structure and semantics. Data retrieval, while providing a result to the user of a database system, does not solve the trouble of retrieving information about topic. To satisfy the user information need, the IR system must understand the contents of the documents in a collection and rank them according to a level of importance to the client query. This interpretation of document content involves extracting syntactic and semantic information from the document text and using this information to match the user information need. The only difficulty is not knowing how to extract this information but also knowing how to use it to decide relevance. Thus, the notion of relevance is at the center of information retrieval. In fact, the primary goal of an IR system is to retrieve all the documents which are relevant to a user query while retrieving as few non-relevant documents as possible.

II. 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.

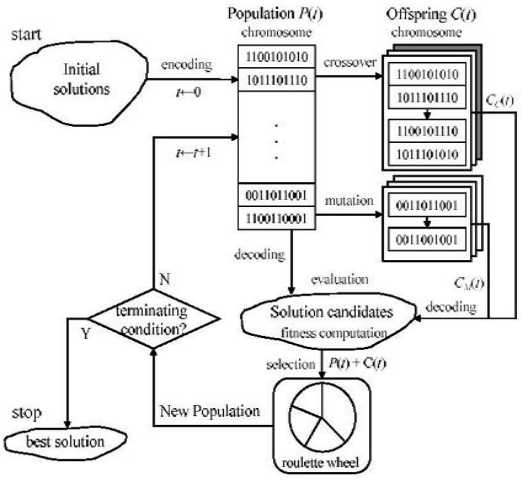

Figure 2. General structure of genetic algorithms

It starts with the operation and the basic theory of the genetic algorithm, and then moves on to the key aspects of current genetic algorithm theory. This theory is centered around the notion of a building block. We will be talking about deception, population sizing studies, the role of parameters and operators, building block mixing, and linkage learning. These studies are motivated by the desire of building better GAs, algorithms that can solve difficult problems quickly, accurately, and reliably. It is therefore a theory that is guided by practical matters.

A. Genetic Algorithm Operation

This section describes the operation of a simple genetic algorithm [2], [3]. The exposition uses a step-oriented style and is written from an application perspective. The steps of applying a GA are:

1. Choose an encoding
2. Choose a fitness function
3. Choose operators
4. Choose parameters
5. Choose initialization method and stopping criteria

B. Encoding

The application of a genetic algorithm to a problem starts with the encoding. The encoding specifies a mapping that transforms a possible solution to the problem into a structure containing a collection of decision variables that are relevant to the problem at hand. A particular solution to the problem can then be represented by a specific assignment of values to the decision variables. The set of all possible solutions is called the search space, and a particular solution represents a point in that search space. In practice, these structures can be represented in various forms, including among others, strings, trees, and graphs. There are also a variety of possible values that can be assigned to the decision variables, including binary, k-ary, real, and permutation values.

Traditionally, genetic algorithms have used mostly string structures containing binary decision variables. We will be assuming this representation for the purpose of illustrating the basic operation and theory of the GA. The terminology used in GAs is borrowed from real genetics. The structure that encodes a solution is called a chromosome or individual. A decision variable is called a gene and its value is called an allele.

C. Fitness Function

Coming up with an encoding is the first thing that a genetic algorithm user has to do. The next step is to specify a function that can assign a score to any possible solution or structure. The score is a numerical value that indicates how well the particular solution solves the problem. Using a biological metaphor, the score is the fitness of the individual solution. It represents how well the individual adapts to the environment. In this case, the environment is the search space. The task of the GA is to discover solutions that have high fitness values among the set of all possible solutions.

D. Operators

Once the encoding and the fitness function are specified, the user has to choose selection and genetic

operators to evolve new solutions to the problem being solved.

The selection operator simulates the "survival-of-the-fittest". There are various mechanisms to implement this operator, and the idea is to give preference to better individuals. Selection replicates individuals with high fitness values and removes individuals with low fitness values.

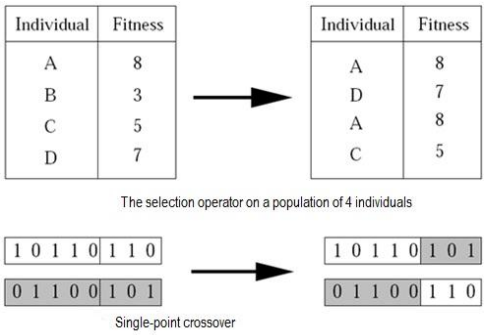

Figure 3. Selection operator for population

Fig. 3 illustrates the selection operator on a population of 4 individuals. If selection was the only operator available, no new solutions would ever be created. In order to explore new solutions, the GA relies on two variation operators: crossover and mutation. Crossover works by pairing members of the population and mixing pieces of one solution with pieces of another solution. The original pair of individuals are called the parents, and the resulting pair of individuals are called the children. This operator is typically applied with a high probability and is responsible for most of the search performed by the GA. Figure 3 illustrates a commonly used crossover operator: single-point crossover. Note however, that there are many types of crossover operators. The key idea is to exchange pieces from one solution with pieces of another solution. Mutation randomly changes the value of a decision variable. With binary coded GAs, this means changing a 0 to a 1 or vice-versa. Within genetic algorithms, mutation is usually considered a background operator that should be used with a very low probability. It is often used as a way to ensure that diversity is never lost at any gene position Parameters with an encoding, a fitness function, and operators in hand, the GA is ready to enter in action. But before doing that, the user has to specify a number of parameters such as population size, selection rate, and operator probabilities.

E. Initialization Method and Stopping Criteria

The last steps of applying a GA are the specification of an initialization method and stopping criteria. The genetic algorithm is usually initialized with a population of random individuals, but sometimes a fraction of the population is initialized with previously known (good) solutions.

Following the initialization step, each individual is evaluated according to the user's specified fitness function. Thereafter, the GA simulates evolution on the artificial population of solutions using operators that mimic the survival-of-the-fittest and principles of natural genetics such as recombination and mutation. The process is repeated a number of times (or generations) until some specified stopping criteria is met. A number of criteria can be chosen for this purpose, including among others, a maximum number of generations or time has elapsed, some predefined fitness value is reached, or the population has converged substantially.

III. LITERATURE SURVEY

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 crossover 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.

$$\text{Recall}=\frac{\text{Number of documents retrieved and relevant}}{\text{Total relevant in collection}}$$

Precision is defined as the proportion of retrieved document that is relevant

$$\text{Precision}=\frac{\text{Number of documents retrieved and relevant}}{\text{Total retrieved}}$$

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. The initial GA system (GA1) uses a measure of cosine similarity between the query vector and the chromosomes of the population as a fitness function, with the following equation:

$$\frac{\sum_{i=1}^{t} x_i . y_i}{\sqrt{\sum_{i=1}^{t} x_i^2 . \sum_{i=1}^{t} y_i^2}}$$

Another GA2 uses a new fitness function represents by following equation:

$$\sum_{i=1}^{t} x_i - y_i$$

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. The algorithm shows the effects of applying GA to get better effectiveness of queries in IR systems. 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. The originality of this work is twofold. First, author proposed new segmentation criteria based on text entailment for interpreting the cohesion and connectivity of segments and second, author used a genetic algorithm which uses a measure based on text entailment for determining the topic boundaries, in order to identify a predefined number of segments. The concept of text entailment is used to indicate the state in which the semantics of a sentence can be inferred from the semantics of another one, i.e. if the fact of an utterance entails the fact of another utterance. Formally speaking, a sentence P (called premise) entails another sentence H (called hypothesis) denoted by $P \rightarrow H$, if and only if H is less informative than P, but P and H have a minimum degree of similarity or, in other words, if similarity of P with respect reported to H is less than the similarity of H with respect to P. The similarities between a pair of sentences were computed using the cosine measure. In order to generate the E matrix, the similarities between all sentence pairs were computed. On the concept of Memetic algorithm (MA), Ziqiang Wang *et. al*. [9] presents 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. Indeed the cumulative total number of relevant documents using MA through all the iterations is higher than using PSO and GA. Therefore, proposed document query optimization algorithm efficiently improves the performance of the query search. Loia and Luengo [10] 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. On the concept of Text Information Extraction based on Genetic Algorithm and Hidden Markov Model. Rong LI *et. al*. [11] present 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. An HMM includes two layers: one observation layer and one hidden layer. The observation layer is the observation sequence for recognition and the hidden layer is a Markov process, (i.e. a limited state machine), in which each state transition all has transition probability. Habiba

et. al. [12] introduces 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 conclude 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 differentiating 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. The idea is to train a classifier to differentiate good expansion terms from others using a number of features associated to the terms and the documents retrieved with the original query, as well as their relationships. Training the classifier requires an appropriate set of training data composed of both, "good" as well as "bad" terms, and distinguished by a number of features that can be extracted from any query and set of documents. If the increased query improves the significance of the retrieved documents the term is judged good and bad otherwise. But the process does not consider any relationship among expansion terms. Proposed method uses a genetic algorithm to generate a set of terms that are appropriate for expansion. The genetic approach has been proven as a strong method to select those terms that maximize the quality of the retrieved documents in terms of MAP. Proposed method concludes that the information extracted from the GA was good enough to improve the overall quality. 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] introduce 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 [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. [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 [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. 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 [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.

IV. Conclusion

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

References

[1] Internet World Stats. Worldwide internet users. (Jan 7, 2013). [Online]. Available: http://www.internetworldstats.com

[2] L. David, *Handbook of Genetic Algorithms*, New York: Van Nostrand Reinhold.,1991,

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

[4] B. Klabbankoh and O. Pinngern, "Applied genetic algorithms in information retrieval," *IJCIM*, vol. 7, no. 3, December 1999.

[5] A. A. Ahmed, *et. al*., "Using genetic algorithm to improve information retrieval systems," *World Academy of Science, Engineering and Technology*, pp. 6-12, 2006.

[6] M. Koorangi and K. Zamanifar, "A distributed agent based Web search using a genetic algorithm," *International Journal of Computer Science and Network Security*, vol. 7, no. 1, pp. 65-76, Jan 2007.

[7] A. Abdelmgeid, "Applying genetic algorithm in query improvement problem," *International Journal "Information Technologies and Knowledge"*, vol. 7, no. 1, pp. 309-316, 2007.

[8] A. Mihaila, A. Mihis, and C. Mihaila, "A Genetic Algorithm for Logical Topic Text Segmentation", in *Proc. Third International Conference on Digital Information Management*, London, 2008, pp. 500-505.

[9] Z. Q. Wang, X. Sun, and D. X. Zhang, "Web document query optimization based on memetic algorithm," *Pacific-Asia Workshop on Computational Intelligence and Industrial Application*., *IEEE*, 2008, pp. 53-56.

[10] A. Veloso, H. M. Almeida, M. Gonçalves, and J. W. Meira, "Learning to rank at query-time using association rules," *SIGIR*, Singapore, 2008, pp. 267-273.

[11] R. Li, J. H. Zheng, and C. Q. Pei , "Text information extraction based on genetic algorithm and hidden markov model," in *Proc. First International Workshop on Education Technology and Computer Science*, 2009, pp. 334-338.

[12] H. Drias, I. Khennak, and A. Boukhedra, "Hybrid genetic algorithm for large scale information retrieval," in *Proc. IEEE International Conference on Intelligent Computing and Intelligent Systems*, Shanghai, 2009, pp. 842-846.

[13] L. Araujo and J. Perez-Iglesias, "Training a classifier for the selection of good query expansion terms with a genetic algorithm," in *Proc. IEEE Conference on Evolutionary Computation*, Barcelona, 2010, pp. 1-8.

[14] P. Bajpai and Dr. M. Kumar, "Genetic algorithm an approach to solve global optimization problems," *Indian Journal of Computer Science and Engineering*, vol. 1, no. 3, pp. 199-206, 2010

[15] A. S. S. Sathya and B. P. Simon, "A document retrieval system with combination terms using genetic algorithm," *International Journal of Computer and Electrical Engineering*, vol. 2, no. 1, pp. 1-6, February 2010

[16] A. Leuski, "Evaluating document clustering for interactive information retrieval.," in *Proc. 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*, vol. 31, 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 *Proc. 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.

Md. Abu Kausar received his BCA degree from T. M. Bhagalpur University, Bhagalpur in 2002, Master in Computer Science from G. B. Pant University of Agriculture & Technology, Pantnagar, India in 2006 and MBA (IT) from Symbiosis, Pune, India in 2012. He has received Microsoft Certified Technology Specialist (MCTS). At present, he is pursuing Ph.D in Computer Science from Jaipur National University, Jaipur, India and he is receiving UGC MANF JRF Fellowship during Ph.D Programme.
He is having 6+ years of experience as a Software Developer. His research interest includes Information Retrieval and Web Crawler.

Md. Nasar received his BCA degree from T. M. Bhagalpur University, Bhagalpur in 2002, Master in Computer Science from G. B. Pant University of Agriculture & Technology, Pantnagar, India in 2006. He has also received Microsoft Certified Technology Specialist (MCTS). At present, he is pursuing Ph.D in Computer Science from Galgotias University, Gr. Noida, India. He is having 6 years of experience in Teaching, and Software Development. His research interest includes Software Reliability and soft computing.

Dr. Sanjeev Kumar Singh is working as Assistant Professor in Department of Mathematics at Galgotias University, Gr. Noida, India. He earned his M. Sc. and Ph.D. degrees with major in Mathematics and minor in Computer Science from G.B. Pant University of Agriculture and Technology, Pantnagar, India. Before that he completed B.Sc. (Physics, Mathematics & Computer Science) from Lucknow University, Lucknow.
He is having more than nine years of teaching and research experience. Besides organizing three national conferences, he has published several research papers in various International/National journals of repute and several national and international conferences and workshops. His areas of interest include Mathematical Modeling, Differential Geometry, Computational Mathematics, data mining & Information Retrieval.