ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

Optimization of Clustered Web Search Queries using Genetic Algorithm for Effective Personalized Web Search

Dr. Suruchi Chawla

Assistant Professor, Dept. of Computer Science, Shaheed Rajguru College of Applied Science for Women, University of Delhi, Vasundhra Enclave , Delhi , INDIA

ABSTRACT: In this paper web search queries are optimized using genetic algorithm based on clustered query sessions for effective Personalized Web Search. The purpose of web queries optimization is to bridge the vocabulary gap between web queries and documents relevant to the queries in order to retrieve relevant documents for effective information retrieval. These optimal set of queries are used for recommendation in order to retrieve relevant documents for personalization of web search according to the information need of the user. Experiment was conducted on the data set of user query sessions captured in three broad domains Academics, Entertainment and Sports and the results confirm the improvement of precision using cluster based optimal queries.

KEYWORDS: Web, Information Retrieval, Search engine, Personalized Web Search, Information Scent, Clustering ,Genetic Algorithm, Optimization, Crossover, Mutation.

I. INTRODUCTION

Information Retrieval(IR) on the web for effective personalization of web search is a big challenge for search engines. The search engines retrieve large collection of search results on the web for a given user input query out of which very few are relevant. It is found that there is a vocabulary gap between the user input query and the keywords used in creating content on the web. Due to this vocabulary gap, the keywords of the user input query could not retrieve relevant web documents early in search results and therefore the precision of search results decreases. A novel approach is proposed in this paper for web queries optimization using genetic algorithm based on clustered query sessions with the objective of bringing the overlap between the vocabulary of user queries and documents on the web for effective information retrieval.

Research has been done for improving the precision of search results in order to better satisfy the information need of the user. [27][52][46][48][3][15][37][40][51] Work related to IR generating related queries recommendations has been done in [17] using Term Query Graph, using phrase substitutions or tools [35][50], based on dynamic knowledge approach in [7], based on clustering process which group semantically similar queries in [23], using session based approach in [6], using snippet based method in [39], based on users behavior information in [54], using query reformulation techniques in [31], using genetic programming for improving the weighted Boolean query formulation in [12], Optimization of Queries in Arabic text Collections using GA in [13], using adaptive genetic algorithm under vector space model in [53]

Thus in this paper novel approach is proposed for web queries optimization using genetic algorithm on clicked documents clustered based on their content. The clustered query sessions has the repository of terms extracted from those documents which were clicked by the users with similar information need. The mutation pool is a set of terms that initially contains terms extracted from the clicked documents of the cluster in a specific domain. Thus the clusterwise repository of terms in mutation pool are more likely to contain the terms relevant to the specific domain and will be effective for user queries optimization using Genetic algorithm so that queries are enriched with domain specific relevant terms.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

An algorithm is proposed for Personalized Web Search based on optimal set of web queries for recommendations. The entire processing of the algorithm is divided into two Phases: Phase I, Phase II. In Phase I, offline processing is performed. During offline processing, user query sessions on the web are processed to generate the query session keyword vectors and clustered. The genetic algorithm is performed on set of queries associated with each cluster in order to generate the optimal queries set. Thus these optimal set of queries are augmented with domain specific terms for retrieving relevant documents early in search results. Thus at the end of processing of Phase I, each cluster is associated with the optimal set of web queries.

In Phase II, online processing is performed. During online processing, the user input query issued for web search is used to select the most similar cluster for generating the recommendations of optimal set of queries along with the search results retrieved using initial input query. The user responses to the current web search results is tracked and stored in the user profile. On the request of next result page, the user profile so far captured is transformed into keyword vector. This user profile keyword vector is then used to select the cluster for generating the next set of recommendation of optimal queries along with the web search result retrieved using the optimal query selected from the previous page. Thus this recommendation of optimal queries along with the personalization of search results using optimal query will soon converge to the information need of the user.

Experiment was conducted on the data set of user query sessions captured on the web in three selected domains Academics, Entertainment and Sports. The performance of the Personalized Web Search using optimal set of queries is compared with PWS using simple queries recommendations (without optimization) in [47]. The results which were verified statistically confirm the significant improvement in precision of search results using the optimal queries.

The subsequent sections of the paper are organized as follows: second section discusses Related Work, third section explains basic concepts required as Background knowledge, fourth section describes the Personalized Web Search using Optimal set of Queries for recommendations, fifth section presents the Experimental Study and the last section concludes the paper.

II. RELATED WORK

Work has been done in research which utilizes Web search engine query logs for mining related queries. In [42] study has been done to make use of Web logs for improving different aspects of search engines. In [25] technique is used for clustering similar queries in order to recommend URLs to frequently asked queries of a search engine. In [32] a method has been used for recommending associated queries based on clustering process over web queries from search engines query log. In [44] web recommender approach is proposed based on web logs analysis and user feedback. The re-ranked list of web pages is recommended to the users by comparing the proposed query with historic pattern. In [54] user behavior information is used for query recommendations. Recommendations are based on snippets clicked by the users and generate good results for low frequency queries.

Query expansion is another approach for recommending related queries. It is found in research that average query terms are near two. [4] It is found that most of the time queries are ambiguous and the promising solution is to expand a query with new terms.

Research has been done on query expansion using number of different existing approaches like works based on query log [59] [38] , general or domain-specific ontology [30][19][16] user profile and collaboratively filtering [28] selecting popular words in the results [18][20][36][55],vector space model in [56] and only term frequency is considered in [55]. Query expansion is also related to the topics of faceted search in [26][1][5], cluster labeling/summarization in [8][33] and result differentiation in [60].

Genetic Algorithm is found to be a powerful search mechanism and is known for its robustness and quick search capabilities. Genetic Algorithm is suitable for information retrieval since the document search space represents a high dimensional space and has been the choice of optimization technique because of the following reasons GA inspired from natural theory of evolution has the ability to work on many solutions in parallel. Due to parallel nature of genetic algorithms it is much more efficient at navigating vast space than traditional algorithms. [29]

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

In this paper novel approach is proposed for Personalization of web search using genetic algorithm for generating optimal queries based on clustered query sessions. The purpose of using genetic algorithm for query optimization is to reformulate the queries using domain specific terms which could retrieve more and more relevant documents early in search results. The effectiveness of the proposed method is further confirmed with experimental results.

III. BACKGROUND

A. *Genetic Algorithm(GA) :*

GA is a search method based on the natural theory of evolution [21]. Genetic Algorithm being insensitive to the initial condition imposed on them discards any solution that is not promising. Due to this feature of Genetic Algorithm it is more flexible, robust and simple in design.[43] Thus the solution space is explored in multiple direction at one time and has a better chance at finding the true global maximum of the system on first try. Thus Genetic Algorithm almost always produces a relatively very good solution to the problem at hand. [2] In [34] the effectiveness of using different Genetic Algorithm with different similarity measure is studied in vector space model based on Arabic data Collection. It is concluded that GA approach which uses one-point crossover operator, point mutation and Inner Product similarity as fitness function is the best IR system in VSM.

During the implementation of GA, the sequence of steps is defined as follows. [9][22][11][10][49]

1. Initialization: In the initialization step, population of chromosomes is initialized using the problem specific domain knowledge. The chromosomes represent the different possible solution to the given problem.
2. Evaluation: After the initialization of the population, the fitness value is defined relative to the problem. The fitness value measures the degree of goodness of the chromosomes in representing the solution to the problem. The selection of population of chromosomes for reproduction in next generations is done on the basis of the fitness value evaluated in this step.
3. Selection: In the selection phase, chromosomes with high fitness values are selected and are allocated more copies in the mating pool for reproduction using recombination operators. There are number of selection methods such as roulette-wheel selection, stochastic universal selection, ranking selection, tournament selection and truncates selection. Tournament Method has been selected in this paper because it yields satisfied solution at earlier generations for the chromosomes representation used in this paper.
4. Recombination: In the Recombination phase, the selected chromosomes are recombined using crossover operator which is a genetic operator for the reproduction of offspring from parent chromosomes. There are various types of crossovers like k-point Crossover, Uniform Crossover, Uniform Order-Based Crossover, Order-Based Crossover, Partially Matched Crossover (PMX). Single point Crossover has been selected for crossover of the chromosomes representation used in this paper.
5. Mutation: In this phase mutation is applied to the selected chromosomes. The mutation is the genetic operator which changes the gene at the specific position in the chromosome. Single point/bitwise mutation is used in this paper since it is commonly used and suited for our problem.
6. Replacement: In the Replacement phase, the offspring population generated using selection, recombination and mutation operators will replace the parent population. Some of them are Delete-all, Steady-state, Steady-state-no-duplicates. The Steady-state-no-duplicates has been used in this paper since it provides the flexibility of specifying the number n of individuals to be replaced and strategy to select these individuals.
7. Steps 2-6 are repeated until a terminating condition is met.

B. *Information Scent :*

Information scent is the sense of value and cost of accessing a page based on perceptual cues with respect to the information need of user. The Inferring User Need by Information Scent (IUNIS) algorithm is used to quantify the Information Scent s_{id} of the pages P_{id} clicked by the user in i^{th} query session. [14][24][41]

In [46] the page access $PF.IPF$ weight and $Time$ are used to quantify the information scent associated with the clicked page in a query session. The information scent s_{id} is calculated for each clicked page P_{id} in a given query session i for all m query sessions identified in query session mining as follows

$$s_{id} = PF.IPF(P_{id}) \times Time(P_{id}) \forall i \in 1..m \; \forall d \in 1..n \quad (1)$$

$$PF.IPF(P_{id}) = \frac{f_{P_{id}}}{\max_{d \in 1..n} f_{P_{id}}} \times \log\left({M}/{m_{P_d}} \right) \quad (2)$$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

$PF.IPF(P_{id})$: PF correspond to the page P_{id} normalized frequency $f_{P_{id}}$ in a given query session i where n is the number of distinct clicked page in session i and IPF correspond to the ratio of total number of query sessions M in the whole data set to the number of query sessions m_{P_d} that contain the given page P_d.

$Time(P_{id})$: It is the ratio of time spent on the page P_{id} in a given session i to the total duration of query session i.

C. Clustering of Query sessions keyword vector

Each query session keyword vector is generated from query session which is represented as follows

query session=(input query,(clicked URLs/Page)$^+$)

where clicked URLs are those URLs which user clicked in the search results of the input query before submitting another query ; '+' indicates only those sessions are considered which have at least one clicked Page associated with the input query.

The query session vector Q_i of the ith session is defined as linear combination of content vector of each clicked page P_{id} scaled by the weight s_{id} which is the information scent associated with the clicked page P_{id} in session i. That is

$$Q_i = \sum_{d=1}^{n} s_{id} * P_{id} \qquad \forall\, i \in 1..m \qquad (3)$$

In the above formula n is the number of distinct clicked pages in the session i and s_{id} (information scent) is calculated for each clicked page present in a given session i as defined in eq 1. The content vector of clicked page P_{id} is weighted using TF.IDF. Each ith query session is obtained as weighted vector Q_i using formula (3). This vector is modeling the information need associated with the ith query session.

The k-means algorithm is used for clustering query sessions keyword vectors since its performance is good for document clustering. [45] [57][58]

The vector space implementation of k-means uses score or criterion function for measuring the quality of resulting clusters. The criterion function is computed on the basis of average similarity between query session keyword vectors and centroid of the assigned clusters.

IV. PERSONALIZED WEB SEARCH USING OPTIMAL SET OF QUERIES FOR RECOMMENDATIONS

In this paper an approach is proposed for Personalization of web search using optimal set of queries for recommendations. The set of queries in a specific domain are optimized using genetic algorithm by adding/replacing with the domain specific terms in order to retrieve more and more relevant documents early in search results. An algorithm is defined in this paper for Personalized Web Search using optimal set of queries for recommendations. The entire processing of the algorithm is divided into two phases: Phase I and Phase II.

In Phase I offline processing is performed. During offline processing, the user query session containing the user input query and associated clicked URLs are transformed into keyword vector using content and Information Scent of clicked URLs. These user query session keyword vectors are clustered using k-means algorithm. Each cluster contains the collection of user query sessions in specific domain and associated clicked URLs which satisfy the similar information need in a given domain. The genetic algorithm is applied on user input queries present in a given cluster for optimization. Thus at the end of offline processing, each cluster is associated with the optimal set of queries. The stepwise execution of Offline processing of Phase I is given below.

Phase I:
Offline Preprocessing
1. Data Set of web queries and the associated clicked URLs is collected on the web and further preprocessed to get the Query Sessions. 2. For each clicked URLs in a given query session, the Information Scent Metric is calculated which is the measure of the relevancy of the clicked URLs with respect to the information need of the user query session. 3. Query sessions keyword vector is generated from query sessions using Information Scent and content of Clicked URLs(TF.IDF) using eq 3. 4. k-means algorithm is used for clustering query sessions keyword vector.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

5. Each cluster j is associated with the mean keyword vector $clust\_mean_j$.
6. For each cluster j maintain the list of Queries in list Q_j.
7. For each cluster j apply the algorithm **Generate Optimal set of queries using genetic algorithm** on the List Q_j associated with the cluster j to determine the optimal set of queries associated with each cluster and is represented by OQ_j (Optimal Queries j).

Algorithm 2:

Generate Optimal set of queries using genetic algorithm

Input: List Q_j, cluster mean keyword vector $clust\_mean_j$

Output: Optimal set of Queries, OQ_j

1. Initial population of chromosomes is generated using the user queries associated with a given $cluster_j$ where each chromosome represents a distinct user query and a gene in a given chromosome represents non zero weight allocated to a particular term in a given user query .The number of genes in each chromosome is equal to the number of distinct terms associated with all user queries in a given cluster.
2. Once the population is initialized with the queries chromosomes, the fitness value of the candidate solution represented by a given chromosome is evaluated. Each chromosome C representing the user query q is evaluated using the Fitness Function. Fitness function Fitness(q) is defined as follows

$$Fitness(q) = \max_{d_i \epsilon A_q} \left(\sigma(t, d_i)\right)$$

where t is the mean keyword vector associated with the cluster and A_q is the set of top 10 url retrieved from search engine using query q. σ is the similarity measure for a document d_i and the mean keyword vector t . Only the snippets of d_i returned by the search engine are used for computing the similarity.

3. Select those chromosomes which have the highest Fitness value using Tournament selection and also followed Elitism which copies the best chromosome (or a few best chromosomes) to new population without mutation and crossover.
4. Apply the single point crossover and single point mutation with mutation probability 0.25 and crossover rate of 0.8 on the selected chromosomes not included in Elitism.
5. Apply the steady-state-no-duplicates replacement policy to replace the population of parent chromosome with the reproduced offspring chromosomes obtained in step 3 and 4 in order to generate the next generation of population P.
6. Goto step 2 until the required number of n1 iterations or terminating conditions is satisfied where the difference between the optimal Fitness values of last 50 generation is less than the threshold value τ.
7. **7.** Upon termination, queries chromosome with maximum fitness value is collected and stored in the set OQ_j associated with the cluster.

In Phase II online processing is performed. During online processing, the user search input query issued for web search is used to select the cluster which is most similar to the information need of the input query. The selected cluster is used to recommend the optimal set of queries along with the search results retrieved from the web using initial input query. The user clicks to the search results are tracked and user selected recommended query is stored in user profile. When the user request for next result page, the user profile containing the clicked URLs is transformed into keyword vector and is used to select the cluster for generating the next set of recommendations of optimal set queries along with the search results retrieved from the web using optimal query selected from the previous page. This process of optimal queries recommendations along with the personalization of search results using the selected optimal query continues till the user information need is satisfied. The stepwise execution of online processing of Phase II is given below.

Phase II

Online Processing.

1. **1.** The search query q1 entered by the user is used to retrieve the search results on the web and at the same time select the j^{th} cluster which is most similar to the information need of the keyword based user input query q1 measured using cosine similarity measure.
2. **2.** The Optimal query set OQ_j associated with the cluster j is selected.
3. **3.** The selected OQ_j is presented to the user along with the search results page retrieved using input query

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

q1.

4. If the user request for the next result page

a. The users clicks to the search results on the current page have been tracked and user selected optimal query are stored in user profile.

b. Model the partial information need of the current user profile using the information scent and content of the URLs clicked so far in his partial user profile and obtain the user session keyword vector $current\_usersessionvector_t$.

c. Select the j^{th} cluster which is most similar to the information need associated with the $current\_usersessionvector_t$

d. The Optimal query set OQ_j associated with the cluster j is selected.

e. The user selected optimal query from the previous page is used to retrieve the web search results for the next requested result page.

f. Thus the optimal queries in the set OQ_j along with the web search results obtained in step e are displayed on the next requested result page.

g. Goto step 4.

else

Current search session is terminated

V. EXPERIMENTAL RESULTS

The experiment was conducted on a data set of web user query sessions collected using the system architecture developed to capture the user's clicks on Google search results. In order to generate the dataset of web query sessions, the user is required to enter the input query through a GUI based interface of the architecture which is then passed on to the Google search engine API to retrieve the search results displayed with the check boxes. The user clicks on the retrieved search results are captured through the check boxes displayed on the GUI and stored in the database.

The captured user query sessions on the web are processed further to find the query session keyword vector using Information Scent and content of clicked URLs. The k-means algorithm is then applied to group the similar information need query session keyword vector in clusters.

The experiment was performed on Pentium IV PC with 120 GB RAM under Windows XP using JSP, JADE , Oracle and genetic algorithm tool box of MATLAB. In the experimental set up for evaluating the performance of personalized web search using optimal set of queries, the values of following parameters are used in the genetic algorithm: MAXGEN, length(P), crossover rate, mutation rate, Tournament Size in the Tournament Selection method. The MAXGEN is the maximum number of generations of population generated in the evolutionary process, length(P) represents the number of chromosomes individuals in the population, crossover rate is the recombination rate of the selected chromosome individuals in the population and mutation rate is the rate of mutating the chromosomes in the population.

In this study, the experiment was conducted with the following values of selected parameters- the size of the population represented as length(P) was m for each cluster where m is the number of user queries in set Q_j associated with each j^{th} cluster, crossover probability was varied in the range of [0.6-0.8] in increment of 0.1 and the mutation rate was varied in the range in [0.1-0.3] in increment of .05.

During offline preprocessing,the clustering agent developed in JADE is executed to generate the clusters of query session keyword vectors. The Genetic algorithm is performed on each cluster in order to get the optimal queries set associated with each cluster. In this study, the process of generating the population continues till the difference in the optimum fitness value of last 50 consecutive generations is less than the threshold value τ=0 .000001.The genetic algorithm tool box of MATLAB software package was used for applying the genetic algorithm on the clustered data set. The experiment was iterated for 100 generation for a given population P and the size of the Tournament in the Tournament Selection was set to 4. The optimal results were obtained at the crossover rate of 0.8 and mutation rate of 0.25.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

The performance of PWS using optimal set of queries is evaluated using the average precision of Personalized Search Results and compared with average precision of Personalized Search Results using simple queries recommendations(without optimization)in [47] in each of the selected domains (Academics, Entertainment and Sports). In order to evaluate the performance, the 25 test queries were selected randomly in each of the domains Academics, Entertainment and Sports. The purpose of selecting the queries in these three domains is to cover wide range of queries on the web. The relevancy of the documents was decided by the experts in the domain to which the queries belong. The recommended optimal set of queries along with the search results are displayed with checkboxes in order to capture user's clicks for the input query 'hindi song' as given in Fig 1 below.

Fig. 1. Shows the Optimal queries recommendations along with the search results displayed with the CheckBoxes to capture the user clicks.

The experimental results in Fig 2 show the average precision of test queries computed in the domains of academics, entertainment and sports using both PWS with optimal queries set/ with queries recommendations(without optimization). The average precision is improved in each of the selected domains using personalized web search with optimal set of queries.

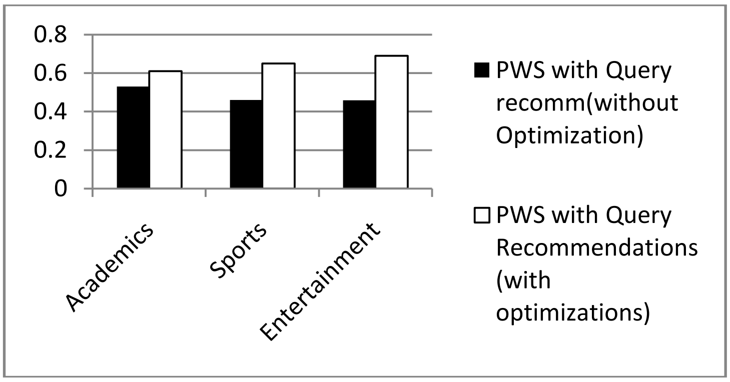

Fig.2. Shows the avgprecision of PWS with Query Recommendations(without/with optimization) in Academics, Sports and Entertainment.

The obtained results were analyzed using the statistical paired t-test for average precision of both PWS with optimal queries set/queries recommendations (without optimization) . The comparison was done on the basis of data set of 25 queries in each of selected domain with 74 degrees of freedom (d.f.) for the combined sample as well as in all three categories (Academics, Entertainment and Sports) with 24 d.f each. It was concluded that average precision improved significantly when personalized web search using optimal set of queries in comparison to improvement in average precision of search results using queries recommendations (without optimization).

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

VI. Conclusion and Future Work

In this paper genetic algorithm is used for the optimization of the clustered web queries in order to bridge the vocabulary gap between user queries and web documents in a specific domain. An algorithm is proposed for Personalized Web Search using optimal set of queries for recommendations. These optimal set of queries contains the term found in the web documents relevant to the user query and can retrieve relevant web pages early in search results. Experiment was conducted on the data set of query sessions captured on the web in the domains: Academics, Entertainment and Sports. The results confirm the improvement in the average precision of search results using Personalized Web Search with cluster based optimal set of queries for recommendations.

References

1. A. Kashyap, V. Hristidis, and M. Petropoulos. 'FACeTOR: Cost-Driven Exploration of Faceted Query Results', in: *Proceedings of the 19th ACM International conference on information and knowledge management CIKM*, pp. 719–728, 2010.
2. Adam Marczyk. 'Genetic Algorithms and Evolutionary Computation.' The Talk.Origins Archive. , 2004,Retrieved March 14, 2014 from the World Wide Web: http://www.talkorigins.org/faqs/genalg /genalg.html
3. B. Arzanian, F. Akhlaghian, , and P. Moradi.' A Multi- Agent Based Personalized Meta-Search Engine Using Automatic Fuzzy Concept Networks', *Third International Conference on Knowledge Discovery and Data Mining*, pp. 208 – 211, 2010.
4. B. J. Jansen, A. Spink, J. Bateman and T. Saracevic. 'Real Life Information Retrieval: a study of user queries on the web'.ACM SIGIR Forum, Vol 32, No.1, pp. 5-17, 1998.
5. C Li, N. Yan, S. B. Roy, L. Lisham, and G. Das. 'Facetedpedia: Dynamic generation of query-dependent faceted interfaces for wikipedia', in: *Proceedings of the 19th international conference on World Wide Web*, pp. 651–660, 2010.
6. C. P. Sumathi, R. Padmaja Valli, and T. Santhanam. 'Automatic recommendation of web pages in web usage mining'. International Journal on Computer science and Engineering (IJCSE). Vol 2 , pp. 3046-3052, 2010.
7. D. Broccolo, O. Frieder, F. Nardini, R. Perego and F. Silvestri . 'Incremental Algorithms for Effective and Efficient Query Recommendation'. SPIRE 2010, LNCS 6393. pp. 13-24, 2010.
8. D. Carmel, H. Roitman and N. Zwerdling. ,'Enhancing Cluster Labeling Using Wikipedia.' in: Proceedings of the 32nd international ACM SIGIR conference on research and development in information retrieval, pp. 139–146, 2009.
9. D. E. Goldberg. 'Genetic Algorithms in Search,Optimization and Machine Learningccluwer Academic Publishers, Norwell, MA, 1989.
10. D. E. Goldberg, and K. Sastry. 'A practical schema theorem for genetic algorithm design and tuning', Genetic and Evolutionary Computation Conference, pp. 328–335, 2001.
11. D. E. Goldberg. 'Design of Innovation: Lessons From and For Competent Genetic Algorithms', Kluwer Academic Publishers, Norwell, MA, 2002.
12. D. H Kraft, F. E. Petry, B. P. Buckles and T. Sadasivan, 'The use of genetic programming to build queries for information retrieval', *Proc. IEEE Symp. Evol. Comput.*, 1994.
13. E.A Mashagba, F. A. Mashagba, and M. O. Nassar. 'Query Optimization Using Genetic Algorithms in the Vector Space Model*'*. International Journal of Computer Science Issues (IJCSI), 8(5), 2011.
14. E H Chi, P. Pirolli , K. Chen and J. Pitkow. 'Using Information Scent to model User Information Needs and Actions on the Web', in: *International Conference on Human Factors in Computing Systems, New York*, NY, USA, pp. 490-497, 2001.
15. F. Akhlaghian, , B. Arzanian, and P. Moradi. 'A Personalized Search Engine Using Ontology-Based Fuzzy Concept Networks', *International Conference on Data Storage and Data Engineering*, pp. 137 – 141, 2010.
16. F. A. Grootjen and ` T. P. van der Weide 'Conceptual Query Expansion.' Data & Knowledge Engineering., Vol 56 No.2, pp.174–193, 2006.
17. Francesco Bonchi, Perego Raffaele, Silvestri Fabrizio, Vahabi Hossein and Venturini Rossano, 'Efficient query recommendations in the long tail via center-piece subgraphs', Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, pp. 345-354, 2012.
18. G. Cao , J.-Y. Nie , J. Gao, and S. Robertson. ' Selecting Good Expansion Terms for Pseudo-Relevance Feedback.' in: *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 243–250, 2008.
19. G Fu., C. B. Jones, and A. I. Abdelmoty. 'Ontology-Based Spatial Query Expansion in Information Retrieval',in *OTM'05: Proceedings of the 2005 OTM Confederated International Conferences on the Move to Meaningful Internet Systems:CoopIS,COA and ODBASE*,Vol Part 2, pp. 1466–1482, 2005.
20. G. Koutrika, Z. M. Zadeh, and H. Garcia-Molina. 'Data Clouds: Summarizing Keyword Search Results over Structured Data', in: *Proceedings of the 12th International Conference on Extending Database Technology :Advances in Database Technology*, pp. 391–402, 2009.
21. H. J. Bremermann. 'The evolution of intelligence. The nervous system as a model of its environment', Technical Report No. 1, Department of Mathematics, University of Washington, Seattle, WA, 1958.
22. H. Holland. 'Adaptation in Natural and Artificial Systems', MIT Press, Cambridge MA,USA, 1992.
23. H. Zahera, G. El Hady, and W. El-Wahed. 'Query Recommendation for Improving Search Engine Results'. World Congress on Engineering and Computer Science (WCECS), San Francisco, USA. Vol. 1. 2010.
24. J. Heer and E.H. Chi. 'Separating the Swarm: Categorization method for user sessions on the web', in: *International Conference on Human Factor in Computing Systems*, pp. 243-250, 2002.
25. J. R. Wen, J. Y. Nie and H. J. Zhang 'Clustering user queries of a Search Engine.' in *W3C: Proceedings at 10th international World Wide Web Conference*, pp 162-168,2001.
26. K. Chakrabarti, S. Chaudhuri, and S. won Hwang. 'Automatic Categorization of Query Results'. in: Proceedings of the 2004 ACM SIGMOD international Conference on management of data, pp. 755–766, 2004.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

27. Kyung-Joong Kim, , and Sung-Bae.Cho . 'A personalized Web search engine using fuzzy concept network with link structure', Joint 9th IFSA World Congress and 20th NAFIPS International Conference, 1, pp. 81 – 86, 2001.
28. L. Fu, D. H.-L Goh and S. S.-B Foo. 'Evaluating the Effectiveness of a Collaborative Querying Environment', *International Conference on Asian Digital Libraries*, pp. 342–351, 2005.
29. L. Tamine, C. Chrisment, and M. Boughanem. 'Multiple query evaluation based on an enhanced genetic algorithm‖', Information Processing and Management , Vol 39 No 2, pp. 215–231, 2003.
30. M. Baziz , M. Boughanem and N. Aussenac-Gilles. 'Conceptual Indexing Based on Document Content Representation'. in *CoLIS: Proceedings of the 5th International Conference on Context:conceptions of library and Information Sciences* , pp.171–186, 2005.
31. M. Boughanem, C. Chrisment, J.Mothe, C. S. Dupuy, and L.Tamine, 'Connectionist and genetic approaches for information retrieval', in *Soft Computing in Information Retrieval: Techniques and Applications*, F. Crestani and G. Pasi, Eds. Heidelberg, Germany: Physica- Verlag, vol. 50, pp. 102–121, 2000.
32. M. Hosseini, and H. Abolhassani. 'Clustering search engines log for query recommendation', Springer-Verlag Berlin Heidelberg , CSICC, CCIS 6, pp. 380-387, 2008.
33. M. Muhr, R. Kern and M. Granitzer. ' Analysis of Structural Relationships for Hierarchical Cluster Labeling.' , in: *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval*, pp. 178–185, 2010.
34. M. O Nassar, F. A. Mashagba and E. A. Mashagba. 'Investigating Genetic algorithms to optimize the user query in the vector space model'. Australian Journal of Basic and Applied Sciences, 7(2): 47-53, 2013.
35. N. Azimi and S. Kiani, 'Accelerating the Response of Query in Semantic Web', International Journal on Computer Network and Information Security, Vol. 8, pp. 26-33, 2014.
36. N. Sarkas, N. Bansal, , G. Das and N. Koudas.' Measure-driven keyword-query expansion', in: *Proceedings of the VLDB Endowment*, Vol 2, No.1, pp.121–132, 2009.
37. O. Nasraoui, and C. Petenes. 'Combining Web Usage Mining and Fuzzy Inference for Website Personalization', in: *International Conference on Knowledge Discovery and Data Mining*, pp.37-46, 2003.
38. P.-A Chirita, C. S. Firan, and W Nejdl.' Personalized Query Expansion for the Web'. In SIGIR'07: *Proceedings of the 30th annual international ACM SIGIR conference on Research and Development in information retrieval*, New York, pp. 7–14, 2007.
39. Poonam Goyal and N. Mehala. 'Concept based query recommendation'. Proceedings of the Ninth Australasian Data Mining Conference-Volume 121. Australian Computer Society. Inc, 2011.
40. P. Navrat, M. Kovacik, A. B. Ezzeddine and V. Rozinajova. 'Web search engine working as a bee hive', Journal Web Intelligence and Agent Systems, 6(4), pp. 441–452, 2008.
41. P. Pirolli. 'The use of proximal information scent to forage for distal content on the world wide web', Working with Technology, Mind: Brunswikian. Resources for Cognitive Science and Engineering, Oxford University Press,2004.
42. R. Baeza-Yates. 'Query Usage Mining in Search Engines.', Web mining: applications and techniques, Anthony scime, Editor, Idea Group, 2004.
43. Richard Baker. 'Genetic Algorithms in Search and Optimization.' Financial Engineering News. 1998, Retrieved December 4, 2004 from the World Wide Web: http://www.fenews.com/fen5/ga.html
44. R. Bhushan and R. Nath . 'Recommendation of optimized web pages to users using Web Log mining techniques'. Advance Computing Conference (IACC), 2013 IEEE 3rd International. IEEE, 2012.
45. R J. Wen, Y J. Nie ,and, J H. Zhang. 'Query Clustering Using User Logs', Journal ACM Transactions on Information Systems, Vol 20 No. 1, pp. 59-81, 2002.
46. S. Chawla, and P. Bedi. 'Personalized Web Search using Information Scent', in: International Joint Conferences on Computer, Information and Systems Sciences, and Engineering, Technically Co-Sponsored by: Institute of Electrical & Electronics Engineers (IEEE), University of Bridgeport, published in LNCS (Springer), pp. 483-488, 2007.
47. S. Chawla and Punam Bedi. 'Improving Information Retrieval Precision by Finding Related Queries with similar Information need using Information Scent.' in *ICETET'08: Proceedings, – The 1st International Conference on Emerging Trends in Engineering and Technology*, (Proceedings published by IEEE Computer Society Press and Papers also available in IEEE Xplore, pp.486-491, July 16-18, 2008.
48. S. Chawla,. 'Personalized Web Search using ACO with Information Scent', International Journal of Knowledge and Web Intelligence, Inderscience, Vol 4, Nos 2/3, 2013.
49. T. B¨ack, D. B. Fogel, and Z. Michalewicz, Handbook of Evolutionary Computation, Oxford University Press, Oxford, 1997.
50. V. Jain and M. Singh, 'Ontology development and query retrieval using protege tool', International Journal of Intelligent Systems and Applications, Vol. 5, No. 9, pp. 6775, 2013.
51. V. Snasel, A. Abraham, S. Owais , J. Platos, and P. Kromer. 'Optimizing Information Retrieval Using Evolutionary Algorithms and Fuzzy Inference System', Foundations of Computational Intelligence,4, pp 299-324, 2009.
52. Wen-Chih Peng and Yu-Chin Lin. 'Ranking Web Search Results from Personalized Perspective', in:*The 8th IEEE International Conference on E-Commerce Technology and The 3rd IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services*, pp.12, 2006.
53. W. Maitah, M. Al-Rababaa and G. Kannan. 'Improving the Effectiveness of Information Retrieval System Using Adaptive Genetic Algorithm'. International Journal of Computer Science & Information Technology, 5(5):91-105, , 2013.
54. Y Liu., Miao Junwei, Zhang Min, Ma Shaoping, and Ru Liyun.' How do users describe their information need: Query recommendation based on snippet click model'. Expert Systems with Applications 38, no. 11 (2011): 13847-13856, 2011.
55. Y. Tao and J. X. Yu. 'Finding Frequent Co-occurring Terms in Relational Keyword Search.' in: *Proceedings of the 12th International Conference on Extending Database Technology:Advances in Database Technology*, pp. 839–850, 2009.
56. Y. Xu, G. J. F. Jones and B. Wang. 'Query Dependent Pseudo-Relevance Feedback based on Wikipedia'. in: *Proceedings of the 32nd international ACM SIGIR conference on research and development in Information retrieval* , pp. 59–66, 2009.
57. Y. Zhao and G. Karypis.' Comparison of agglomerative and partitional document clustering algorithms', SIAM Workshop on Clustering High-dimensional Data and its Applications, 2002a.
58. Y. Zhao, , and G. Karypis. 'Criterion functions for document clustering: Experiments and Analysis'. Technical report,University of Minnesota, Minneapolis, MN, 2002b.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

59. Z. Bar-Yossef and M Gurevich.' Mining Search Engine Query Logs via Suggestion Sampling'. PVLDB, 1(1), pp.54–65, 2008.
60. Z. Liu, P. Sun and Y. Chen. 'Structured Search Result Differentiation', in: *Proceedings of the VLDB Endowment*, Vol 2 No.1, pp.313–324, 2009.

BIOGRAPHY

Dr. Suruchi Chawla is a Assistant Professor in Shaheed Rajguru College of Applied Science for Women. University of Delhi, INDIA. She received her Phd in Computer Science in 2011 from University of Delhi, India. Her research interests are Information Retrieval, Search Engines, Personalization of Web Search, Web Mining etc.