

# Domainwise Web Page Optimization Based On Clustered Query Sessions Using Hybrid Of Trust And ACO For Effective Information Retrieval

Dr. Suruchi Chawla

**Abstract:** In this paper hybrid of Ant Colony Optimization(ACO) and trust has been used for domainwise web page optimization in clustered query sessions for effective Information retrieval. The trust of the web page identifies its degree of relevance in satisfying specific information need of the user. The trusted web pages when optimized using pheromone updates in ACO will identify the trusted colonies of web pages which will be relevant to users information need in a given domain. Hence in this paper the hybrid of Trust and ACO has been used on clustered query sessions for identifying more and more relevant number of documents in a given domain in order to better satisfy the information need of the user. Experiment was conducted on the data set of web query sessions to test the effectiveness of the proposed approach in selected three domains Academics, Entertainment and Sports and the results confirm the improvement in the precision of search results.

**Index Terms:** Information Retrieval, Search engines, Personalized Web Search, Clustering, Information Scent, Trust, Ant Colony Optimization.

## INTRODUCTION

Information on the web is huge in size and growing continuously hence it is a big challenge to identify the relevant web pages for a specific information need of the user. Research has been done for better personalizing the web search of the user using various techniques. [4][15] [20] [69] [14] [63] [12] [34] [68] [64] [50][49] [38] [51] [52] [53][54] [55][37]. It is found that hybrid of optimization technique Ant Colony Optimization (ACO) and trust have been applied successfully in various domains and results proved promising. In [65] ACO algorithm is adopted in trust model in order to simulate trust relationship between cloud entities in cloud computing. In [22] trust evaluation model is developed in P2P networks. In [5] ant algorithms have been applied to estimate the degree of trust between nodes in mobile adhoc networks. In [56] ACO algorithm has been used for reaching optimality in grid scheduling problem. In [10] a hybrid of ant colony system for predicting the recommendations in trust based recommender system is proposed which considers all the target item ratings along the solution path rather than just stopping and using the first rating found in the search process. In [59] a optimized and trusted routing in MANET is proposed where nodes with trust value above the threshold value are considered for routing and finally optimization is done using ACO to yield more optimized performance. It is found that hybrid of ACO and trust shows promising results in various domains but it is realized in this research that benefit of this hybrid can be used in query log mining for web page optimization in a given domain in order to identify relevant documents for effective web information retrieval. Thus in this paper hybrid of ACO and trust has been proposed for web page optimization using clustered query sessions. The significance of using hybrid of Trust with ACO is that use of trusted web pages for optimization will increase the quality of colonies of web pages identified using ACO and will identify

the web pages relevant to the information need of the user in a given domain. These trusted colonies of web pages in a specific domain when selected for recommendations will retrieve more and more relevant documents and improve the precision of search results. Thus user response to trusted web pages is captured to update the trust and pheromone of clusters and is used for identifying more and more trusted colonies of web pages for further recommendations. This process of recommendations and optimization continues till the web search is personalized to the information need of the user. Since the ratio of relevant documents to retrieved documents is increased hence web search will soon converge to information need of the user effectively. The flowchart of the proposed work is shown below in Fig 1.

Experiment was conducted on the data set of web search query sessions captured in three domains Academics, Entertainment and Sports and the results confirms the improvement in the precision of search results. The rest of sections are organized as follows. section 2 discusses the related work, section 3 describes the background required for understanding the proposed approach, section 4 explains the proposed approach of Domainwise web page optimization using hybrid of trust and ACO for effective Information Retrieval, section 5 presents the experimental study to analyze the effectiveness proposed approach and section 6 concludes the paper.

- *Dr. Suruchi Chawla is Assistant Professor in Shaheed Rajguru College of Applied Science for Women, University of Delhi, INDIA, PH-9810342692. E-mail: sur\_chawla@rediffmail.com*



**Fig 1.** Shows the flowchart of the proposed approach of Domainwise web page optimization using trust and ACO .

## 2. RELATED WORK

It is found that recommender system can be more effective by incorporating trust than traditional collaborative filtering. [41][46][32][9][60]. In [40] trust based recommender system is proposed using both trust metric and similarity metric. In [44] trust propagation model is proposed using number of hops in trust propagation and the trust value is calculated between source to destination participant. In [36] a method is proposed for generating recommendations by trusted friends only. In [16] a novel trust model is developed based on recommendations

in online service oriented environment. ACO is a nature inspired metaheuristic for the solution of hard combinatorial optimization (CO) problems. A metaheuristic is a general-purpose algorithmic framework that defines the heuristic methods which can be applied to different optimization problems with relatively few modifications. [27][29][28][30] ACO has been used to solve many optimisation problems such as sequential ordering [21], scheduling [6], assembly line balancing [7], probabilistic travelling salesman problem (TSP) [35], DNA sequencing [8], 2D-HP protein folding [3] and protein-ligand docking [33]. In [31] Ant Colony Optimization (ACO) is applied to build query association graphs from the query logs for the purpose of query recommendation. In [48] clustering based on ants is used to access to a variety of collaborative learning agent groups so as to fully mobilize the enthusiasm of collaborative learning team members. In [62] the ant colony optimization is applied on the log data to build an adaptive domain model automatically in order to satisfy user's information request effectively in more structured collections such as digital libraries, local Web sites, and intranets. In [39] combination of ant based clustering and fuzzy c-means is proposed. In [57] model is proposed which combines the ACO and Fuzzy logic to generate the list of recommendations to online users based on the comparison of the user's navigational behavior with other user's data. In [47] an ACO algorithm is developed called Ant-Recommender System which sense pheromones found on their clusters in order to determine the best cluster for recommending the items with in clusters of user profiles. In [36] recommender system is proposed based on collaborative behavior of ants by integrating trust between users. The hybrid of ACO and trust has been applied in various domains and shows promising results. It is found no work has been done which uses the hybrid of ACO and trust for effective web page optimization based on query log mining. The advantage of using hybrid of ACO and trust in query log mining for web page optimization is that it generates recommendations of relevant web pages using ACO based optimization on trusted colonies of web pages. Thus in this paper a novel approach is proposed using ACO and trust in query log mining for effective personalized web search. It is found in [53] that the performance of the Personalized Web Search was improved when converted into optimization problem and solved using Ant Colony optimization techniques by replacing the pheromone in ACO with Information Scent. In this paper work in [53] has been extended to apply the hybrid of trust and ACO in query log mining for identifying trusted colonies of web pages for recommendations in order to personalize the web search of the user more effectively.

## 3 BACKGROUND

### 3.1 Trust

The concept of Trust has been gaining increase amount of attention in research communities like online recommender system. A trust is defined as social phenomena and the model of trust for artificial world like web is based on how trust works between people in society.[1] Although vast literature on trust has grown in various areas of research with varying meaning of trust but a complete formal unambiguous definition of trust exists rarely in the literature.[11] In [61] the general properties of trust in e-services were surveyed and analyzed and the general properties of trust are listed as follows:

- Trust is relevant to specific transactions only.
- Trust is a measurable belief.
- Trust is directed.
- Trust exists in time.
- Trust evolves in time, even within the same transaction..
- Trust between collectives does not necessarily distribute to trust between their members.
- Trust is reflexive,
- Trust is a subjective belief.

In [18] the “trust” is defined as the reliability of a partner profile to deliver accurate recommendations in the past. Two models of trust called profile and item level are described for generating reliable and accurate recommendations. Thus this trust has been incorporated into collaborative recommendation process and hence generates trust-based weighting and trust-based filtering, both of which can be used with either profile-level or item-level trust metrics. It is found that use of trust values has the positive impact on the overall prediction error rate thus significantly reducing the prediction error rate and thereby improves the prediction accuracy.

### 3.2 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. Information Scent is measure of degree of relevancy of clicked URLs with respect to the information need of the user. [42][43]

#### 3.2.1 Information Scent metric

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 [13][17]. 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

$$sid = PF.IPf(P_{id}) * Time(P_{id}) \forall i \in 1..m \forall d \in 1..n \quad (1)$$

$$PF.IPf(P_{id}) = f_{P_{id}} / \max(f_{P_{id}}) * \log(M/m_{P_{id}}) \quad (2)$$

$d \in 1..n$

$PF.IPf(P_{id})$ :  $PF$  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  $m_{P_{id}}$  that contain the given page  $P_{id}$ .

$Time(P_{id})$  : It is the ratio of time spent on the page  $P_{id}$  in a given session  $i$  to the total duration of query session  $i$ . [49]

#### 3.2.2 Generation of Query sessions keyword vector for Clustering

Each query session keyword vector is generated from query session containing the clicked URLs where each clicked URLs are those URLs which user clicked in the search results of the input query before submitting another query. The query session keyword vector  $Q_i$  of the  $i^{th}$  session is defined as linear combination of TF.IDF 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 formula

(3). This vector is modeling the information need associated with the  $i^{th}$  query session.

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

The k-means algorithm is used for clustering resulting query sessions keyword vectors and the criterion function is used for measuring the quality of resulting clusters. [45][66][67]

### 3.3 Ant Colony Optimization

ACO is inspired from the social behavior of ant colonies as the ants colonies forage for food starting from their nest to the food source using pheromone lying on the paths to the food source. Ants communicate with each other indirectly through chemical called pheromone released by the ants on their path to food source. [23][24][2] In ACO, artificial ants construct the solution to problem represented in the form of construction graph  $GC(V, E)$ , where  $V$  is a set of vertices and  $E$  is a set of edges. Ants move from vertex to vertex along the path while constructing solution to the problem. The algorithm for ACO is given below [58]

Algorithm 1: Basic flow of ACO [25]

1. Represent the solution space by a construction graph.
2. Set ACO parameters and initialize pheromone trails
3. Generate ant solutions from each ant's walk on the construction graph and update pheromone trails.
4. Update pheromone intensities.
5. Go to step 3, and repeat until convergence or termination condition is reached.

As per the flow of ACO algorithm, the objective of ACO's third step is to construct the path of the ant from source to destination by making local optimal choice at each point of decision using probability rule. The probability rule makes use of pheromone trail intensities and heuristic information. The paths constructed by the ants are then used in next step to update the pheromone trail on the paths visited by the ants. [19]

$$P_{ij}^k(t) = \frac{[\tau_{ij}(t)]^\alpha * [\eta_{ij}]^\beta}{\sum_{l \in N_i^k} [\tau_{il}(t)]^\alpha * [\eta_{il}]^\beta} \quad , j \in N_i^k$$

$$P_{ij}^k(t) = 0, j \notin N_i^k \quad (4)$$

- Where:
- $P_{ij}^k(t)$  is the probability of the  $k^{th}$  ant to move from node  $i$  to node  $j$  at the  $t^{th}$  iteration/time step.
- $N_i^k$  is the set of nodes in the neighborhood of the  $k^{th}$  ant in the  $i^{th}$  node.

$P_{ij}^k(t) = 0, j \notin N_i^k$  means the ants are not allowed to move to any node not in their neighborhood. The neighborhood definition is problem-specific.  $[\tau_{ij}(t)]^\alpha$  is the pheromone amount on the arc connecting node  $i$  and node  $j$ , weighted by  $\alpha$ .  $[\eta_{ij}(t)]^\beta$  is the heuristic value of the arc connecting node  $i$  and  $j$ , weighted by  $\beta$ .  $\alpha$  and  $\beta$  are weight parameters that control the relative importance of the pheromone versus heuristic information. The fourth step of ACO algorithm performs the updation of pheromone trails. In the beginning of the optimization, the pheromone value of all arcs on the constructed path is initialized to the constant value  $\tau_0$ . The pheromone trails are updated in two ways as defined in eq 5 and eq 6. Initially pheromone trails on the paths decreases due to evaporation.

poration factor of pheromone where evaporation rate is usually set to be sufficiently fast in order to favor new area of search space and avoid the premature convergence to local optimum. Secondly the pheromone trail on the arcs of the visited path are increased directly proportional to quality of the path. This pheromone updation process increases the probability of the quality paths to be followed by the more ant in future while finding solution to the problem. The quality paths include solution component that were either used by the many ants in the past or was at least followed by the ant which produces high quality solution[19][26]

$$\tau_{ij}(t+1) \leftarrow (1-\rho) * \tau_{ij}(t) + \sum_{k=1}^m \Delta\tau_{ij}^k(t) \quad \forall i, j \in A, 0 \leq \rho < 1 \quad (5)$$

$$\Delta\tau_{ij}^k(t) = Q/C^k(t), (i, j) \in T^k(t) \quad (6)$$

$$\Delta\tau_{ij}^k(t) = 0, \text{ otherwise}$$

Where Q is an application-specific constant, m is the number of ants, A represents all arcs of the problem of construction graph,  $C^k(t)$  is the overall cost function of tour  $T^k(t)$  constructed by the kth ant at the tth iteration, and  $T^k(t)$  is the set of all arcs visited by ant k at the iteration t. Other variations of ACO, however, restrict pheromone depositing to the arcs of the best tour  $T^{\text{best}}$  only.

#### 4 DOMAIN WISE OPTIMIZATION OF WEB PAGES USING HYBRID OF TRUST AND ACO FOR EFFECTIVE PERSONALIZATION OF WEB SEARCH.

In this paper an approach is proposed for clusterwise optimization of trusted clicked web pages using ACO for effective personalized web search. The trust metric is used to identify those web pages which have been clicked most of the time and is relevant to the information need of the user. These trusted web pages are selected for recommendations and the user's response to recommended trusted web pages are captured along with the updation of pheromone value in order to optimize the relevant web pages in a given domain. The use of trusted web pages for optimization increases the quality of web pages colonies identified using ACO based on pheromone value as the updation of pheromone value corresponding to user's clicks to web pages will be more reliable since only trusted web pages are selected for recommendations to the user. Thus the precision of recommended search results is improved as the ratio of relevant documents to the total retrieved documents is increased. The processing of proposed approach is divided into two phases Phase I and Phase II. In Phase I offline processing is performed and In Phase II online processing is performed. During offline processing, the query sessions on the web are transformed into keyword vector using information scent and content of clicked URLs. The query session keyword vectors are clustered using k-means algorithm. Each cluster is associated with clicked URLs which satisfy similar information need on the web. The clicked URLs in a given cluster is associated with the trust and pheromone value where information scent is used to initialize its pheromone value and trust is initialized to zero. Each cluster is associated with trust defined status which is initially false indicating that trust is initially undefined and it becomes true when cluster is selected for recommendation.

#### Offline Preprocessing

1. Data Set Collected on the Web is preprocessed to get the Query Sessions.
2. For each clicked URLs in the query session, the Information Scent Metric is calculated using Eq. (1)
3. Query sessions keyword vector is generated from query sessions using Information Scent and content of Clicked URLs using Eq. (3).
4. k-means algorithm is used for clustering query sessions keyword vector.
5. Each cluster i is associated with the mean keyword vector  $\text{cluster}_i\text{\_mean}$ .
6. For each cluster i the initial value of pheromone is calculated as follows
  - a. Each clicked URLs of the user query session is associated with the initial pheromone value  $\tau_{\text{pheromoneclickedURLs}}(0)$  = Information Scent of the Clicked URLs.  $\Delta\tau_{\text{URLs}} = 0$ . where  $\Delta\tau_{\text{URLs}}$  is the quantity trail substance (pheromone in real ants) laid on Clicked URLs by the kth user/ant between time t and t+n;
  - b. For each distinct clicked URL in the given cluster identify  $\tau_{\text{avgpheromoneURLs}}(0)$  which is calculated over all the query sessions present in the given cluster.
7. Clicked count and recommended count are defined for each distinct clicked URLs and are initialized to zero in the list L associated with each cluster
8. Initialize Trust(ClickedURLi)=0 for each distinct clicked URL in the List L associated to each cluster i.
9. For each cluster i the initially the trust is undefined InitialTrustDefined (i)=false

During online processing, the users are assumed as ants that have some information need to be satisfied. The user's clicks to the recommended trusted web pages are used to update the pheromone of clicked URLs with the information scent of the clicked pages. The pheromone of clicked web page is dynamic which will be increased or decreased with time depending on the number of clicks it receives and for how much time it remain unclicked. At the same time trust values of clicked urls and cluster is updated. Thus the relevancy of the clicked URLs measured using trust and pheromone is not static but changes dynamically depending on user response to the recommended clicked URL. Thus the optimization of trusted clicked URLs using ACO recommend more and more relevant web pages for a specific information need and personalizes the user search effectively.

#### Online Processing

1. The input query is used to find the most similar cluster.
2. For each cluster i the similarity is measured using the formulae
 
$$\text{MatchScore}_i(\text{input query}, \text{cluster}_i) = \text{sim}(\text{input query}, \text{cluster}_i\text{\_mean})$$
3. Identify the most matching cluster i.
4. if for the selected cluster i the InitialTrustDefined(i)=true then



- a. Identify the clicked URLs in selected cluster  $i$  whose Trust value  $\text{Trust}(\text{ClickedURL}_i) \geq \varepsilon$  and store it in list  $L$ .
- Else if  $\text{MatchScore}_i(\text{input query}, \text{cluster}_i) > 0.5$  and  $\text{initialtrustdefined}(i) = \text{false}$ 
  - i.  $\text{InitialTrustDefined}(i) = \text{true}$
  - ii. Identify the clicked URLs in selected cluster  $i$  whose  $\text{avgpheromone} \geq \text{threshold}(\rho)$  and store it in list  $L$ 
    - Else
      - i. Msg "No Trusted Recommendations. Terminate the current session, Initiate new search session"
    - ii. Goto step 1.

5. The list  $L$  is presented to the user in decreasing order of their pheromone value.
6. The Recommendedcount of the recommended URLs in list  $L$  are incremented by 1.
7. The user response to the recommended URLs is tracked and stores it in current user profile.
8. For each recommended URL clicked by the user, the ClickedCount of the corresponding recommended URLs is incremented by 1.
9. The trust value of the recommended URLs are updated as given below.
 
$$\text{Trust}(\text{URL}_i) = \{1 - \text{Distrust}(\text{URL}_i)\} \times \text{Recommendedcount}(\text{URL}_i) / \text{ClickedCount}(\text{URL}_i)$$

$$\text{Distrust}(\text{URL}_i) = \left\{ \frac{\text{Recommendedcount}(\text{URL}_i) - \text{ClickedCount}(\text{URL}_i)}{\text{Recommendedcount}(\text{URL}_i)} \right\}$$
10. For each clicked URL of the current user session if present in the selected clusters, then pheromone value of the clicked URLs w.r.t current user profile will be added to the average pheromone value of the corresponding clicked URLs in selected cluster  $i$  in order to update its average pheromone value.

$$\begin{aligned} \tau_{\text{avgpheromoneURLs}}_{ij}(t) &= (1 - \rho) \\ &\times \tau_{\text{avgpheromoneURLs}}_{ij}(t-1) \\ &+ \tau_j(t) \end{aligned}$$

where  $j$  is the URLs clicked by the current user at time  $t$  and present in the selected cluster  $i$ . This is for all  $j$  where  $j$  is the clicked URLs of the current user and  $j \in$  selected cluster  $i$  and  $\tau_j(t)$  is the information scent of clicked URL  $j$  in the current user session.  $\tau_{\text{avgpheromoneURLs}}_{ij}(t)$  is the average pheromone of the  $j^{\text{th}}$  clicked URL in  $i^{\text{th}}$  cluster.

11. If the user request for the next result page
  - a. 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  $\text{current\_usersessionvector}_t$ .
  - b. The similarity is measured for each  $i^{\text{th}}$  cluster using the formulae
 
$$\text{MatchScore}_i(\text{cluster}_i, \text{current\_usersessionvector}_t) = \text{sim}(\text{current\_usersessionvector}_t, \text{cluster}_i\text{\_mean})$$
12. Identify the most matching cluster  $i$ .

13. Goto step 4.

End

## 5 EXPERIMENTAL STUDY

Experiment was conducted on the data set of user query sessions collected on the web. The architecture is developed using JADE, JSP and database Oracle to generate the data set of users query sessions on the Google search results and performs the personalization of web search based on clustered query sessions. In order to capture the dataset of user query sessions on the web, the user is required to enter the input query through a GUI based interface of the architecture. This input query is passed on to the Google search engine API, and the search results are retrieved and displayed along with the check boxes on the user interface. The Fig 2 shows the search results for input query "hindi songs". The user clicks, on the retrieved search results, are captured through the check boxes displayed on the GUI and are stored in the database.

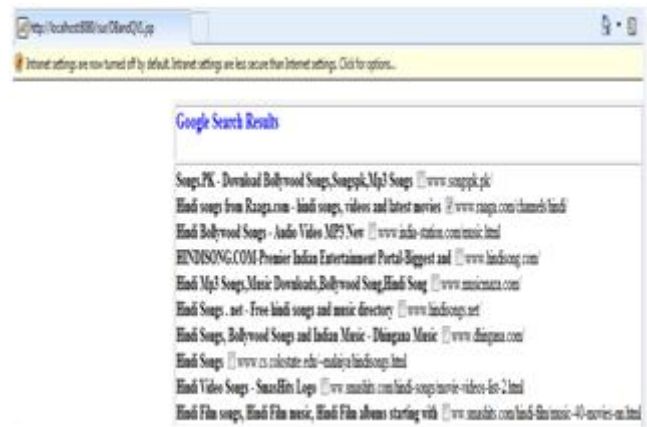
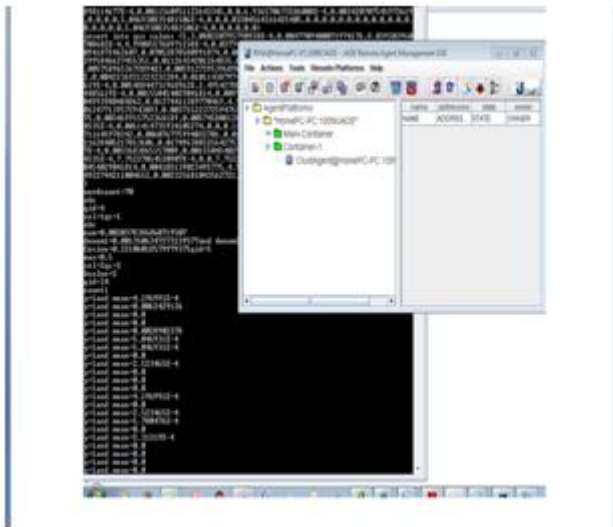


Fig. 2. Shows the Google search results for the input query "hindi songs".

The experiment was performed on duo core processor with 2 GB RAM under Windows XP using JSP, JADE and Oracle. In the experimental set up for evaluating the performance of Personalized Web Search using ACO and Trust, the following parameters set as follows. The values of the threshold value of Trust  $\varepsilon$  is set to 0.5, the pheromone evaporation factor  $\rho = 0.5$  as the experiments was conducted with different value of parameters based on the preprocessed collected data set, where the value of  $\rho \in [0, 1)$  (Dorigo, 1992), and Information Scent lies in  $[0, 1]$ . It is further found by setting the value of  $\rho$  below 0.5, the evaporation rate of the pheromone is not effective enough to the capture the changing user's need in the clustered query sessions, therefore the pheromone evaporation factor is set at  $\rho = 0.5$ . The Agents are developed in JADE to perform the Offline processing associated with proposed approach. The tf.idf vector of the clicked URLs of the query sessions are fetched using the Web Sphinx Crawler and loaded into database using oraloader. The clustering agent developed in the JADE is executed to generate query session keyword vector using Information Scent and clustered them using k-means algorithm. It performs the initialization of the trust of the clusters and the clicked URLs of the query sessions. The pheromone value  $\tau$  of each clicked URLs in the query sessions is initialized to its Information Scent. Snap Shots of the execution of Clustering Agent is shown in Fig 3.



**Fig. 3.** Shows the execution of clustering agent clustering the query sessions keyword vectors.

During Online processing, the input query is issued to GUI based interface designed each for both Personalized Web Search with ACO and trust/with ACO/with Trust only based on the same clustered query sessions dataset. In Personalized Web Search with Trust and ACO, the input query is used to find the cluster most similar to the information need of the current user. The resultant set of the high trusted clicked URLs associated with the selected cluster are recommended and displayed in the GUI Interface as shown in Fig 4.



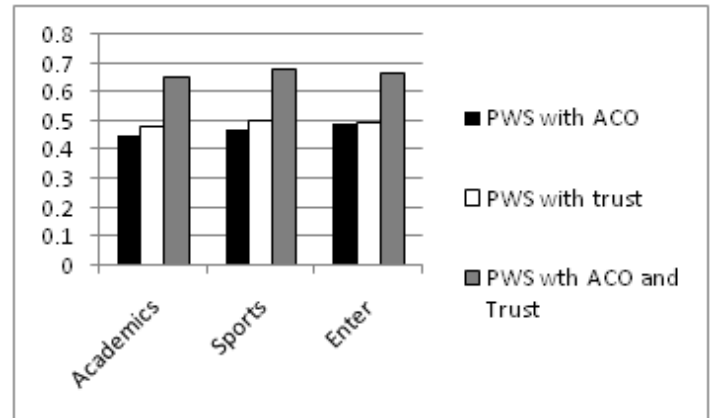
**Fig. 4.** Shows the Personalized Search Results using ACO and trust

The users clicks to the personalized search results are tracked to capture the user's profile and dynamically update the pheromone and trust associated with the stored clicked URLs. This dynamic updation of the pheromone of the trusted clicked URLs using ACO optimizes relevant set of clicked URLs for each identified information need associated with the clusters.

### 5.1 Results & Discussion

The performance of the Personalized Web Search using ACO with Trust is evaluated from the average precision of personalized search results generated by the proposed approach and compared with Personalized Search Results with ACO proposed in [53] and trust in [51]. In order to evaluate the perfor-

mance, the test queries were chosen in three domain Academics, Entertainment, Sports. The number of test queries in the three domains was 25 in Academics, 25 in Entertainment and 25 queries in Sports. During online searching, these test queries were issued in each of the selected domain to the GUI based interface to retrieve the personalized search results with ACO and trust/with ACO/with trust. The average precision is computed by determining the relevant documents retrieved in the personalized search results.



**Fig. 5.** Compares the average precision of personalized search results in Academics, Entertainment and Sports domain using ACO/Trust /ACO and Trust.

The experimental results are shown in Fig 5. The obtained results were also analyzed using statistical paired t-test for average precision of PWS( ACO and Trust) with PWS(with ACO /with trust ) with 74 degrees of freedom (d.f.) for a combined sample as well as for all three categories (Academics, Entertainment and Sports) with 24 d.f, 24 d.f and 24 d.f. The observed value of t for average precision of PWS(ACO and trust) with PWS( ACO) was 57.53 and with PWS(trust) was 44.48 for a combined sample. Value of t for paired difference of average precision of PWS(ACO and trust) versus PWS(with ACO ) and PWS(with trust) was 48.03 and 20.03 for academics, 35.97 and 40.44 for entertainment , 39.22 and 25.98 for sports categories . It was observed that the computed t value for paired difference of average precision lie outside the 95% confidence interval in each case. Hence Null hypothesis was rejected and alternate hypothesis was accepted in each case and it was concluded that average precision is improved significantly using proposed Personalized Web Search with ACO and Trust due to increase in the number of relevant documents to the retrieved documents. Thus the improvement in the precision of search results shows the effectiveness of PWS with ACO and Trust in comparison to PWS( with ACO only/with trust only).

### 6 CONCLUSION AND FUTURE WORK

In this research an approach is proposed for personalized web search using Trust and ACO. The use of trusted web pages for optimization identifies identifies the colony of web pages frequently accessed by the users with similar information need. Thus the optimization of trusted clicked urls will retrieve more and more relevant documents and improves the precision of search results. Experiment was conducted on the data set of web query sessions collected in the three domains entertainment, academics and sports. The results show the significant

improvement in the precision of search results and personalize the web search effectively by catering to the information need of the user. In future this work can be extended to reduce the dimensionality of clustered query sessions in order to reduce the computational complexity and further improve the performance of the proposed system.

## REFERENCES

- [1] A Abdul-Rahman, and S. Hailes. "Supporting trust in virtual communities". Proceedings of the 35th Hawaii International Conference on System Sciences, Hawaii, HI, 2000.
- [2] A Colomi, M. Dorigo, V. Maniezzo, and M. Trubian. "Ant System for Job-shop Scheduling. Belgian Journal of Operations Research", Statistics and Computer Science, Vol. 34 No.1,39-53,1994.
- [3] A Shmygelska, and H. H. Hoos. "An ant colony optimisation algorithm for the 2D and 3D hydrophobic polar protein folding problem", BMC Bioinformatics, Vol. 6 No.30, 2005.
- [4] B Arzianian, F. Akhlaghian, and P. Moradi. "A Multi-Agent Based Personalized Meta-Search Engine Using Automatic Fuzzy Concept Networks." In Third International Conference on Knowledge Discovery and Data Mining, WKDD'10 (pp. 208-211). IEEE, 2010, January.
- [5] B. Wang, X. Chen AND W. Chang, "A light-weight trust-based QoS routing algorithm for ad hoc networks". Science Direct, 2013.
- [6] C. Blum. "Beam-ACO—Hybridizing ant colony optimization with beam search: An application to open shop scheduling", Computers and Operations Research, Vol. 32 No.6, 1565–1591, 2005.
- [7] C. Blum. "Beam-ACO for simple assembly line balancing". INFORMS Journal on Computing, Vol. 20 No.4, 618–627, 2008.
- [8] C. Blum, M. Yabar and M. J Blesa. "An ant colony optimization algorithm for DNA sequencing by hybridization". Computers and Operations Research, Vol. 35 No.11, 3620–3635, 2008.
- [9] C. Hwang, and Y. Chen. "Using trust in collaborative filtering recommendation". Lecture Notes in Computer Science, Volume 4570, 1052-1060. Innovation Network (2009), 2007.
- [10] D. Alathel, "Ant colony inspired models for trust-based recommendations "(Doctoral dissertation, The George Washington University), 2015.
- [11] D. H. McKnight, and N. L. Chervany. "What Trust Means in e-Commerce Customer Relationships: An interdisciplinary conceptual typology". International Journal of Electronic Commerce, 6(2), 35-59, 2002.
- [12] D. Zhou, S Lawless,. and V. Wade. "Improving search via personalized query expansion using social media". Information Retrieval, 1-25, 2012.
- [13] E H. C hi, P. Pirolli, K. Chen and J. Pitkow. "Using Information Scent to model User Information Needs and Actions on the Web" in ACM CHI 2001: Proceedings of the Conference on Human Factors in Computing Systems, New York,NY, USA, 490-497, 2001.
- [14] F. Liu, C. Yu, and W. Meng, "Personalized web search for improving retrieval effectiveness." IEEE transactions on Knowledge and Data Engineering, 16(1), 28-40, 2004.
- [15] H. J. Kim, S. Lee, B. Lee, and S. Kang. "Building concept network-based user profile for personalized web search." IEEE/ACIS 9th International Conference in Computer and Information Science (ICIS), 567-572, IEEE, 2010, August.
- [16] J. Golbeck and J. Hendler, "Inferring Trust Relationships in Web-Based Social Networks", ACM Trans. Internet Technology, vol. 6, no. 4, pp. 497–529, 2006.
- [17] J. Heer, and E.H Chi. "Separating the Swarm: Categorization method for User Access Session on the Web" in ACM CHI 2002: Proceedings of Conference on Human Factor in Computing System, 243-250, . 2002.
- [18] J. O'Donovan, and B. Smyth. "Trust in Recommender Systems." Proceedings of the 10th International Conference on Intelligent User Interfaces. 167-174, 2005.
- [19] K. Jones and A. Bouffet. "Comparison of Ant Colony Optimisation and Differential Evolution" in CompSysTech'07: Proceedings of the International Conference on Computer Systems and Technologies, 2007.
- [20] K. T. Leung, W. Ng, D. L. Lee. "Personalized concept-based clustering of search engine queries.", IEEE Transactions on Knowledge and Data Engineering, 20(11), 1505-1518, 2008.
- [21] L. M. Gambardella and M. Dorigo. "Ant colony system hybridized with a new local search for the sequential ordering problem". INFORMS Journal on Computing, Vol.12 No.3, 237–255,2000.
- [22] L. Yang, L. Qin, Z., Wang, C., and Liu, Y. "A P2P reputation model based on Ant Colony Algorithm". International Conference on Communications, Circuits and Systems, 236-240, 2010.
- [23] M. Dorigo, V. Maniezzo, and A. Colomi. "Positive feedback as a search strategy", Tech. Report 91-016, Dipartimento di Elettronica, Politecnico di Milano, Italy,1991.
- [24] M. Dorigo. "Optimization, learning and natural algorithms (in Italian)", Ph.D. Thesis, Dipartimento diElettronica, Politecnico di Milano, Italy,1992.
- [25] M. Dorigo, and T. Stützle. "Ant Colony Optimization". MIT Press, Cambridge. ISBN: 978-0-262-04219-2, 2004.
- [26] M. Dorigo, and K. Socha. "An Introduction to Ant Colony Optimization". In T. F. Gonzalez (ed.), Approximation Algorithms and Metaheuristics, CRC Press, 2007.



- [27] M. Dorigo, V. Maniezzo, and A. Colnari. "Ant System: Optimization by a colony of cooperating agents", *IEEE Transactions on Systems, Man, and Cybernetics—Part B*, Vol 26 No.1, pp. 29–41, 1996.
- [28] M. Dorigo, and G. Di Caro. "The Ant Colony Optimization meta-heuristic, in *New Ideas in Optimization*", D. Corne et al., Eds. (pp 11-32). London, UK: McGraw Hill, 1999.
- [29] M. Dorigo, G. Di Caro, and L.M. Gambardella. "Ant algorithms for discrete optimization", *Artificial Life*, Vol.5 No. 2, pp. 137–172, 1999.
- [30] M. Dorigo, and T. Stützle. "Ant Colony Optimization." MIT Press, Cambridge. ISBN: 978-0-262-04219-2, 2004.
- [31] M.-D. Albakour, U. Kruschwitz, N. Nanas, D. Song, M. Fasli, and A. De Roeck. "Exploring Ant colony optimisation for adaptive interactive search" in *ICTIR'11: Proceedings of the International Conference on the theory of Information Retrieval*, September, Bertinoro, Italy, 2011.
- [32] N. Lathia, S. Hailes, and L. Capra, 2008. "Trust-based collaborative filtering". *Proceedings of the joint iTrust and PST Conference on Privacy, Trust Management and Security*. Springer, 119-134.
- [33] O. Korb, T. Stützle, and T. E. Exner. "An ant colony optimization approach to flexible protein ligand docking." *Swarm Intelligence*, Vol.1 No.2, pp.115–134, 2007.
- [34] P.-A. Chirita, C. S. Firan, & W. Nejdl. "Personalized Query Expansion for the Web." *30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007)*. Amsterdam, The Netherlands: ACM, 7-14, 2007.
- [35] P. Balaprakash, M. Birattari, T. Stützle, Z. Yuan, and M. Dorigo. "Estimation-based ant colony optimization algorithms for the probabilistic travelling salesman problem". *Swarm Intelligence*, Vol. 3 No.3, 223–242, 2009.
- [36] P. Bedi, and R. Sharma. "Trust based recommender system using ant colony for trust computation". *Expert Systems with Applications*, 39(1): 11831190, Tarrytown, NY, USA, 2012.
- [37] P. Bedi, H. Kaur and S. Marwaha, "TrustBased Recommender System for Semantic Web", *Proc. Int'l Joint Conf. Artificial Intelligence (IJCAI)*, pp. 2677–2682, 2007.
- [38] P. Bedi and S. Chawla. "Agent based information retrieval system using information scent". *Journal of Artificial Intelligence*, 3(4), 220-238, 2010.
- [39] P. M. Kanade, and L. O. Hall. "Fuzzy Ants as a Clustering Concept." In *Proc. of the 22nd Int. Conf. of the North American Fuzzy Information Processing Soc.*, 227–232, 2003.
- [40] P. Massa, and B. Bhattacharjee. "Using trust in recommender systems: An experimental analysis." *Proceedings of the Second International Conference on Trust Management*, Oxford, UK., 221-235, 2004.
- [41] P. Massa, and P. Avesani. "Trust-aware Recommender Systems." *Proceedings of the ACM Conference on Recommender Systems*, 17-24, 2007.
- [42] P. Pirolli. "Computational models of information scent-following in a very large browsable text collection", in *ACM CHI 97: Proceedings of the Conference on Human Factors in Computing Systems*, 3-10, 1997.
- [43] P. Pirolli. "The use of proximal information scent to forage for distal content on the world wide web." In *Working with Technology in Mind: Brunswikian. Resources for Cognitive Science and Engineering*, Oxford University Press, 2004.
- [44] R. Guha, R. Kumar, P. Raghavan and A. Tomkins, "Propagation of Trust and Distrust," *Proc. Int'l Conf. World Wide Web (WWW'04)*, pp. 403–412, 2004.
- [45] R. J. Wen, Y. J. Nie, and J. H. Zhang. "Query Clustering Using User Logs." *ACM Transactions on Information Systems*, 20(1), 59-81, 2002.
- [46] R. Levien. "Attack-resistant Trust Metrics." Ph.D. thesis, University of California at Berkeley, USA, 2004.
- [47] R. Sharma, Singh, M., Makkar, R., Kaur, H., AND Bedi, P. 2007. "Ant Recommender: Recommender system inspired by ant colony", in *Proceedings of International Conference on Advances in Computer Vision and Information Technology*, 361-369.
- [48] Rui Zeng and Ying-yan Wang. "Research of personalized Web-based intelligent collaborative learning, *Journal of Software*", Vol.7 No.4, 904-912, 2012.
- [49] 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 Bridgport, published in LNCS (Springer), 483-488, 2007.
- [50] S. Chawla and P. Bedi. "Improving information retrieval precision by finding related queries with similar information need using information scent." In *First International Conference on Emerging Trends in Engineering and Technology, ICETET'08*, 486-491. IEEE, 2008.
- [51] S. Chawla. "Trust in Personalized Web Search based on Clustered Query Sessions." *International Journal of Computer Applications*, 59(7), 36-44, 2012a.
- [52] S. Chawla. "Semantic Query Expansion using Cluster Based Domain Ontologies." *International Journal of Information Retrieval Research (IJIRR)*, 2(2), 13-28, 2012b.
- [53] S. Chawla. "Personalised web search using ACO with information scent." *International Journal of Knowledge and Web Intelligence*, 4(2), 238-259, 2013.



- [54] S. Chawla. "Personalised Web Search using Trust based Hubs and Authorities." *International Journal of Engineering Research and Applications*, 4(7), 157-170, 2014a.
- [55] S. Chawla. "Novel Approach to Query Expansion using Genetic Algorithm on Clustered Query Sessions for Effective Personalized Web Search ". *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(11), 73-81, 2014b.
- [56] S. Kumar , and M. Singh. "Adaptive and dynamic load balancing in grid using ant colony optimization." *International Journal of Engineering and Technology*. 4(4): 167-174, 2012.
- [57] S. Nadi, M. H. Saraee, M. D. Jazi, and A. Bagheri. "FARS: Fuzzy Ant based Recommender System for Web Users", *International Journal of Computer Science Issues*, 8(1), 2011.
- [58] S. Sorlin , C. Solnon and J.-M. Jolion. "A Generic Graph Distance Measure Based on Multivalent Matchings" in *SCI: Studies in Computational Intelligence*, 52, Springer, pp. 151–182,2007.
- [59] S. Sridhar, and R. Baskaran. "ANT Based Trustworthy Routing in Mobile Ad Hoc Networks Spotlighting Quality of Service." *Open Journal of Computer Science and Information Technology*, 3(1), 064-073, 2015.
- [60] T. Peng, and T. Seng-cho. "iTrustU: A blog recommender system based on trust and collaborative filtering." *Proceedings of the ACM Symposium on Applied Computing*. New York, NY. 1278-1285,2009.
- [61] T. Dimitrakos. "A Service-Oriented Trust Management Framework." *International Workshop on Deception, Fraud & Trust in Agent Societies*, 53-72, 2003.
- [62] U. Kruschwitz, M.-D. Albakour, J. Niu, J. Leveling, N. Nanas, Y. Kim, D. Song, M. Fasli and A. DeRoeck. "Moving towards Adaptive Search in Digital Libraries". In *Advanced Language Technologies for Digital Libraries*, 2011.
- [63] W. C. Peng, and Y. C. Lin, (2006, June). "Ranking Web search results from personalized perspective." *The 8th IEEE International Conference on E-Commerce Technology and The 3rd IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services* (pp. 12-12)., 2006,IEEE.
- [64] W .Gao, C. Niu, J.-Y. Nie, D. Zhou, J. Hu, K.-F. Wong, & H.-W. Hon. "Cross-Lingual Query Suggestion Using Query Logs of Different Languages." *30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007)*. Amsterdam, The Netherlands: ACM, 463-470, 2007.
- [65] W. Lin, and S. Alvarez. "Efficient adaptive-support association rule mining for recommender systems." *Data Mining and Knowledge Discovery Journal*, 6(1), 2004.
- [66] Y. Zhao, and G. Karypis. "Comparison of agglomerative and partitional document clustering algorithms." In *SIAM Workshop on Clustering Highdimensional Data and its Applications*, 2002a.
- [67] Y. Zhao, and Y. Karypis. "Criterion functions for document clustering." *Technical report, University of Minnesota, Minneapolis, MN*, 2002b.
- [68] Z. Yin, M. Shokouhi, and N. Craswell. "Query Expansion Using External Evidence." *Lecture Notes In Computer Science. 31st European Conference on Information Retrieval (ECIR 2009)*. Toulouse, France: Springer, pp. 362374, 2009.
- [69] Z. Zhu, J. Xu, X Ren, Y.Tian, and L. Li, (2007, October). "Query Expansion Based on a Personalized Web Search Model." *Third International Conference In Semantics, Knowledge and Grid*, 128-133,2007, IEEE.