

Personalized Web Search Using Trust Based Hubs And Authorities

Dr. Suruchi Chawla

(Department of Computer Science, Shaheed Rajguru College of Applied Science, University of Delhi, Delhi-96, INDIA)

ABSTRACT

In this paper method has been proposed to improve the precision of Personalized Web Search (PWS) using Trust based Hubs and Authorities(HA) where Hubs are the high quality resource pages and Authorities are the high quality content pages in the specific topic generated using Hyperlink- Induced Topic Search (HITS). The Trust is used in HITS for increasing the reliability of HITS in identifying the good hubs and authorities for effective web search and overcome the problem of topic drift found in HITS. Experimental Study was conducted on the data set of web query sessions to test the effectiveness of PWS with Trust based HA in domains Academics, Entertainment and Sport. The experimental results were compared on the basis of improvement in average precision using PWS with HA (with/without Trust). The results verified statistically show the significant improvement in precision using PWS with HA (with Trust).

Keywords: Information Retrieval, Search engine, Web, Information Scent, Clustering, Trust, Hubs, Authorities, HITS, Personalized Web Search.

I. INTRODUCTION

Information Retrieval on the web relevant to the need of the user is a big challenge. Search engine retrieve a large collection of documents from the web for a given query out of which very few are relevant and are further ranked lower because of insufficient keyword used in the user input query. During web search, users rarely go beyond the first results page and relevant documents are ranked lower in the search results, hence the precision of search results decreases. Personalized Web Search aims at improving the precision of search results by customizing the web search according to the information need of the user by bringing more and more relevant documents higher in search results. Extensive work has been done in an area of personalized web search for improving the precision of search results. [20][29][62][18][49][40][52] Various page ranking methods for Personalized Web Search have been proposed in research in order to rank more and more relevant documents higher in search results. The high ranking of relevant documents improves the precision and hence satisfies the information need of the user effectively. Each method has its advantages and some limitation. In [26] web page ranking is calculated by computing hub and authorities score of the pages but it has the limitation of topic drift and efficiency. The research in this paper has been focused in improving the quality of hubs and authorities for effective Personalized Web Search. An approach is proposed in this paper for Personalized Web Search using trust based hubs and authorities where the trust of web

pages is used in HITS for generating the Hubs and authorities. The trust increases the reliability of HITS in identifying the good hubs and authorities and overcome the problem of topic drift in HITS. There is no issue of efficiency found in computing the trust based hubs and authorities since Hubs and authorities have been computed during offline processing of web search.

Work related to the proposed approach has been done in [50] where HITS algorithm without using Trust has been applied on the clustered query sessions. Each cluster is associated with High Information Scent Clicked URLs. Information Scent is the quantitative measure of the relevancy of the clicked URLs in the user query session with respect to the information need of the user. The high scent clicked URLs of each cluster is used as root set for HITS to generate the hubs and authorities for a given cluster. During the web search, the user search input query is used to select the cluster most similar to the information need of the user input query and the selected cluster is used to generate the hubs and authorities for recommendations. The problem of efficiency found in HITS in [26] for computing Hubs and Authorities is overcome in [50] but it is found that pages pointing to and pointed by the high scent web pages in a given cluster have the deviation from the topic of a cluster and hence the topic drift problem still exists. It is found that Information Scent based Hubs and Authorities identifies hubs and authorities on the basis of usage statistics of clicked URLs in a given session with respect to the entire data set but not on the measure of how reliable the

clicked URL was in satisfying the information need of the user in a given topic. Reliability of clicked URL measure how often a given recommended clicked URL is actually clicked by the users during web search in a given topic. Trust is the measure of reliability of recommended clicked URL in satisfying the information need of the user. Hence this research is motivated to use the trusted web pages in a given topic for generating its trust based Hubs and Authorities using HITS. The use of high Trust web pages in HITS will increase the reliability of HITS in identifying good hubs and authorities in a specific topic due to the reasons that trusted web pages are more likely to point to and pointed by high quality pages in a given topic. Hence with these initial set of trusted web pages in a specific topic, the trust based hubs and authorities are identified using HITS for effective Personalized Web Search. Thus the problem of topic drift(deviation from topic) is overcome using trusted web pages in HITS.

In this paper an algorithm is proposed for PWS with Trust based Hubs and Authorities using clustered query sessions. Trust based Hubs and Authorities associated with each cluster are used for recommendations during the Personalization of web search. The entire processing of the algorithm proposed for PWS using trust based hubs and authorities is divided in two phases: Offline and Online.

The offline processing is the processing done before search input queries are processed for Personalized Web Search. During Offline processing, clusters of query sessions are generated where each query session is defined as the user input query and associated clicked URLs. In order to cluster query sessions, the query session keyword vectors are generated from query sessions using Information Scent and content of clicked URLs (TF.IDF). The clustered query sessions is composed of the clicked URLs satisfying the similar information need. Each clicked URL is associated with the trust score which is measure of the reliability of clicked URL in satisfying the information need associated with the clusters in which it is present. The high trust value clicked URL associated with the each cluster are selected using trust threshold value and used as Root set for HITS algorithm. Root set of each cluster is expanded to Base Set which includes all web pages link to and linked by the web pages in the root set upto depth d . The set of web pages in the base set B form the Web Graph $G=(V,E)$ where V is set of set of web pages and E is the set of links connecting the web pages. Trust of the web pages in the graph G is transferred to other web pages using trust propagation method such as logarithm splitting and maximum Share. The trust attenuation is used to limit the flow of trust from root set of trusted pages through an

outlinks to the child nodes as number of links traversed from the root set increases.

Thus the trust propagation methods along with trust attenuation generate the trust based web graph G for each cluster of query sessions where each node in G is associated with trust value. The HITS algorithm is applied on this trust based web graph $G=(V,E)$. In order to apply the HITS on trust based web graph G , trust value of each node in G is used to initialize its hub y_p and authority score x_p . The trust based hub and authority score of each node in Graph G is updated using HITS algorithm. Upon the termination of HITS algorithm on Graph G , each node is associated with two trust based score : hub(y_p) and authority(x_p). The node having high value of x_p than y_p is categorized as good authority than hub and vice versa. Thus each cluster is associated with the Trust based Hubs and Authorities which are further selected using the threshold value set for Trust.

During online processing the input query is used to find the cluster similar to the information need of the user. The selected cluster is used to recommend the trust based Hubs and authorities ranked in decreasing order of their trust value. The user responses to the recommended HAs are captured in profile and are further used to update the trust of recommended HAs and the selected cluster. The user's clicks so far captured in his profile is used to infer his partial information need and is used to select the cluster for the recommendation of Trusted Hubs and Authorities for the next result page. This recommendation process and updating of trust value continues till the search is personalized to the Information need of the user.

Experimental Study was conducted on the data set of query sessions captured on the web in three domains viz academics, entertainment and sports to test the effectiveness of trust based hubs and authorities for personalized web search. The experimental results of the PWS with Trust based Hubs and Authorities were compared with PWS with HA (without Trust). The results verified statistically confirm the significant improvement in the precision of search results using trust based hubs and authorities.

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 Trust based Hubs and Authorities, fifth section presents the Experimental Study and the last section concludes the paper.

II. RELATED WORK

In [33] the misearch system is developed which improves search accuracy by creating user profiles from their query histories and/or examined search

results. These profiles are used to re-rank the results returned by an external search service by giving more importance to the documents related to topics contained in their user profile. In [7] a new approach is proposed for the search named Just-in-Time IR (JITIR) where the information system proactively suggests information based on a person's working context, automatically identifying their information needs and retrieving useful documents without requiring any action by the user. Google labs released an enhanced version of Personalized Search that builds the user profile by means of implicit feedback techniques, adapts the results according to needs of each user, assigning a higher score to the resources related to what the user has seen in the past [36]. In [5] Wifs (Web Information Filtering System) is described which evaluates and reorders page links returned by the search engine, taking into account the user model who typed in the query. Compass filter in [4] follows a similar collaborative approach, but it is based on Web communities by analyzing the Web hyperlink structure, similarly to the HITS algorithm. [26]

In [56] ranking of Web search results is proposed from personalized perspective. In this common access patterns from user browsing activities are mined to automatically obtain user interests. According to the user interests mined and feedbacks of users, a new approach is proposed with the plan of dynamically altering the ranking scores of Web pages. In [48] graph based algorithm based on link structure of web pages is used for web page ranking. The back links are used for rank calculation. The rank is calculated on the basis of the importance of pages. The results computed at the indexing time not at the query time are considered one of its limitations.

In [14] a popularity-based search engine used a popularity-based search algorithm, ranking URLs in order of popularity, with the pages visited most by other users ranking highest in their search results. Outride Inc., an information retrieval technology company acquired by Google (2001), introduced a contextual computing system for the personalization of search engine results. [21]

In [13] HITS algorithm is introduced to identify the Hubs and Authorities in a specific topic relevant to the query. Hubs are the pages linked to many relevant authoritative pages (e.g. link lists for certain topics). Authorities are pages that are referenced by many hubs. It returns pages of high relevancy and importance but it has the limitation of less efficiency and topic drift. In [54] a web page ranking algorithm is proposed which probabilistically estimates that clear semantics and the identified authoritative documents corresponds better to human intuition. It efficiently provides answer to quantitative bibliometric questions. In this method number of factors has to be decided prior and there is the risk of

getting stuck in local maxima. In [57] the page rank is calculated on the basis of weight of the page with the consideration of the outgoing links, incoming links and title tag of the page at the time of searching. It gives higher accuracy in terms of ranking because it uses the content of the pages but it is based only on the popularity of the web page. In [45] algorithm ranks the page by providing different weights based on three attributes i.e. relative position in page, tag where link is contained and length of anchor text. This method has less efficiency with reference to precision of the search engine. The obtained relative position was not so effective, indicating that the logical position not always matches the physical position. In [28] adjacency matrix is used which is constructed from agent to object link not by page to page link. Three vectors i.e. hub, authority and reputation are needed for score calculation of the blog. It is specifically suited for blog ranking. In [6] the ranking of web pages is based on reinforcement learning which consider the logarithmic distance between the pages. This algorithm consider real user by which pages can be found very quickly with high quality. It has the limitation that large calculation for distance vector is needed, if new page is inserted between the two pages. In [19] Visitor time is used for ranking. In this method sequential clicking is used for sequence vector calculation with the use of random surfing model. It is useful when two pages have the same link structure but different content. In [53] the algorithm used for ranking is based on the analysis of tag on social annotation web. This method produces exact ranking results however co-occurrence factor of tag is not considered which may influence the weight of the tag. In [17] it provides the ranking algorithm for semantic search engines. The algorithm uses information extracted from the queries of the user and annotated resources. In this ranking algorithm every page is to be annotated with respect to some ontology which is the tough task. In [30] individual models are generated from training queries. A new query search results are ranked according to the combined weighted score of these models. In this method limited numbers of characteristics are used to calculate the similarity. In [34] popular items are suggested for tagging. In this method three randomized algorithms are used i.e. frequency proportional sampling, move to set and frequency move to set. Tag popularity has been boost up because large number of tag is suggested by this method. In this method alternative user choice model, alternative rule for ranking and alternative suggestive rules are not considered. In [35] page ranking is done using score fusion techniques. It is used when two pages have same ranking. In [58] moving objects are retrieved in uncertain database using Prank(Probabilistic ranked query) and J-Prank(Probabilistic ranked query join). Experimental

results are promising only with limited number of parameters.

In [50] HITS is applied on the clustered query sessions to generate the hubs and authorities ranked using Information Scent for Personalized Web Search. Once the clusters are associated with the High Scent Hubs and Authorities, the user search query is used to select the cluster most similar to the information need of the user input query. The selected cluster is used to recommend the associated High Scent Hubs and Authorities ranked in decreasing value of Information Scent for the Personalization of web search of user. In [47] survey was done on Hyper link Induced Topic Search Algorithms for Web Mining it is found that HITS has topic drift problem.

In [51] trust has been introduced for Personalized Web Search based on clustered query sessions where trust is defined both at the clicked URLs and cluster level. Trust is the measure of the reliability of the clicked URLs and cluster in generating the recommendations relevant to the information need of users. The experimental results show the improvement in the precision due to the use of trust in personalization of the user web search. It is realized in this research that the effectiveness of HITS in identifying the good hubs and authorities can be improved if HITS uses the trusted web pages in a given topic for computing the Hubs and authorities. The use of Trust in HITS reduces the topic drift and increases the reliability of HITS in identifying the good hubs and authorities in a specific topic. The use of trust based hubs and authorities for Personalized Web Search can lead to effective improvement in the precision of search results in comparison to the Personalization of Web Search using Hubs and Authorities(without trust) in [50]. Hence the effectiveness of PWS in satisfying the information need of the user is increased using Trust based Hubs and Authorities.

Extensive Research has been done in the area of trust. In [11] it is demonstrated that positive relation between trust and user similarity holds on the basis of data from current trust based recommender systems. It is shown that difference in the rating of movies decreases as the trust in the reviewing user increases. Research has been done in recent years based on trust-based recommendation in [32][31][10][37][24][39]. In the research it is found that there are purely trust based recommender system, hybrid recommender system and integrated approach. In purely trust based recommender system only, recommendations are done only on the basis of trust. In hybrid approach, the trust based recommendation is used as complementary to other recommendation techniques. In integrated approach trust information is integrated in other recommendation techniques. The approaches based on pure trust based

recommender techniques are proposed in [37][24][39]. In [32] trust based recommendation is used in combination with content based filtering. First in content based filtering, the similarity of the item to be recommended and the item that were previously used or bought is computed using the features of the items. If the clear vote for or against the item could not be provided by the content based filtering then the recommendations will based on the experience of trusted users.

In [27] an integrated approach is developed by enhancing collaborative filtering with the use of trust information directly in the standard prediction formula of GroupLens. The results shows that all trust based approaches improve the accuracy of recommendations. On the basis of recommendation type, trust-based filtering and trust-weighting Reviews are distinguished. In trust based filtering, the information is filtered on the basis of the trustworthiness of the users providing them. In [37] a similar approach is taken in Moleskiing in which ski tour descriptions provided by trustworthy peers are shown to a user. In [23] [42] [38] approaches used are based on trust networks or social networks respectively for spam filtering.

In Trust-weighting, the recommendation for an item is based on the reviews on this item which are then weighted with the trust in the users providing the reviews. The trust weighted reviews are aggregated. The Film Trust website generates the trust based movie recommendations based on Trust-weighting. [24][22]

In this paper research has been motivated to use trust in HITS for effective PWS by generating the trust based hubs and authorities using clustered query session. The HITS is applied on trust based web graph generated in a given topic using trust propagation and attenuation method in order to identify the trust based hubs and authorities. The drawback of topic drift in HITS while computing Hubs and Authorities is overcome with the use of trusted web pages in a given topic. Hence for an effective personalization of web search of the user an algorithm is proposed in this paper for PWS using trust based hubs and authorities.

III. BACKGROUND

3.1 Trust

The concept of Trust has been gaining increase amount of attention in research communities like online recommender system. Trust has been defined and used in many different ways. 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. [2] 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. [15]

One of the definitions of trust given by Dasgupta is “the expectation of one person about the actions of others that affects the first persons choice, when an action must be taken before the actions of others are known” [41]. In another definition given by [12] it is quoted as “trust is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action and in a context in which it affects his own action”. In [9] trust is stated as “trust as the expectation of other persons goodwill and benign intent, implying that in certain situations those persons will place the interests of others before their own”. In [3] two general definition of trust is given, one is called reliability trust also called as evaluation trust and other is decision trust. Evaluation trust can be interpreted as the reliability of something or somebody. It can be defined as the subjective probability by which an individual, A, expects that another individual, B, performs a given action on which its welfare depends. On the other hand, the decision trust captures broader concept of trust. It can be defined as the extent to which one party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible. In [1] two categories of the trust is defined one is Context-specific interpersonal Trust and second is system / impersonal trust. In Context-specific interpersonal Trust, user trust another user with respect to one specific situation but not necessarily another. In system/impersonal trust, user trust in a system as a whole.

Characteristics

In [55] 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.

It is found that trust-based recommendations outperformed collaborative filtering algorithms in certain cases. In [27] 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

3.1.1 Trust Propagation Methods

Trust is propagated among web pages in the web graph using Trust propagation methods. There are two approaches for Trust propagation one is splitting which includes the methods for distributing the trust score from parent to its children and other is trust accumulation which includes the methods of accumulating the trust scores received from the inlinks on the given child node.

In splitting there are three methods:

Equal Splitting : For a node i with $O(i)$ outgoing links and trust score $TR(i)$ will give $d*TR(i)/O(i)$ share to each of its child. d is a constant with $0 < d < 1$.

Constant Splitting: For a node i with trust score $TR(i)$ will give $d * TR(i)$ to each child.

Logarithm Splitting: For a node i with $O(i)$ outgoing links and trust score $TR(i)$ will give $d*TR(i)/\log(1+O(i))$ to each child.

The term d is the decay factor, which determines how much of the parents' score is propagated to its children.

In the accumulation step, There are three method.

Simple Summation: In this the trust values is added from each of the child's Parent to the given child.

Maximum Share: In this maximum trust value among the parent's trust score is propagated to its child.

Maximum Parent: In this sum of the those parents trust values is propagated to a given child in such a way so that it never exceed the trust score of the most trusted parent. [8]

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. The users on the web tend to click those pages in the retrieved search results on the web which seem to satisfy the user's information need. More the page is satisfying the information need of user, more will be the information scent perceived by the user associated to it and more is the probability that the page is clicked by the user. The interactions between user need, user action and content of web can be used to infer information need from a pattern of surfing. [43][44]

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. [16][25]

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

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

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

$$d \in 1..n$$

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

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 i^{th} 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

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

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 i^{th} query session is obtained as weighted vector Q_i using formula (3). This vector is modeling the information need associated with the i^{th} query session.

3.2.3 Clustering of Query session keyword vector

The k-means algorithm is used for clustering query sessions keyword vectors since its performance is good for document clustering. [46][59]

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 vectors and centroid of the assigned clusters. The criterion function I is defined as follows:

$$I = 1/M \sum_{p=1}^k \sum_{v_i \in C_p} sim(v_i, c_p) \quad (4)$$

where C_p be a cluster found in a k-way clustering process ($p \in 1..k$), c_p is the centroid of p^{th} cluster, v_i is the vector representing some query session belonging to the cluster C_p and M is the total number of query sessions in all clusters as defined below. [60]

$$M = \sum_{p=1}^k |C_p| \quad (5)$$

The centroid c_p of the cluster C_p is defined as below:

$$c_p = (\sum_{v_i \in C_p} v_i) / |C_p| \quad (6)$$

where $|C_p|$ denotes the number of query sessions in cluster C_p and $sim(v_i, c_p)$ is calculated using cosine measure.

IV. Personalized Web Search using Trust based Hubs and Authorities.

In this paper an algorithm is proposed for personalized web search using trust based hubs and authorities recommendations based on clustered query sessions. This method is based on using clustered query sessions where each cluster is associated with the Trust based hubs and authorities. In order to generate the trust based hubs and authorities for a given cluster, Trust value is associated with both cluster and clicked URLs. Trust of a cluster is the measure of the goodness of a cluster in generating the reliable recommendations in the past during the personalization of the user web search. Each trusted cluster is associated with the list of trusted clicked URLs where trust of the clicked URLs is a measure of how often the recommended clicked URL of the cluster was clicked by the user in a given topic during the search session. The high trust value clicked URLs associated with each cluster is used in HITS algorithm for generating trust based hubs and authorities. Thus an subalgorithm is proposed for generating trust based hubs and authorities associated with a given cluster using HITS. The processing of this subalgorithm is divided into two part A and B. In part A, the high trust clicked URLs associated with a given cluster form the Root set. Root set are further crawled on the Web using its outlink and inlink to collect the total set of web pages including the root set in the base set B. The web graph $G=(V,E)$ is formed using base set B where V is the set of vertices representing the web pages and E represents the link between the web pages in the base set B for a given cluster.

The trust is propagated in the web graph G associated with a given cluster using Logarithm Splitting and Maximum Share method. In logarithmic

splitting, for a given parent node i and trust value $Trust(i)$, the trust propagated to its children is $Trust(i)/\log(1+O(i))$ where $O(i)$ is the number of outlinks of the parent node i in the web graph. Maximum share method is applied to accumulate the maximum trust on a given child node i from one of its parents through the inlink. The logarithm splitting and Maximum Share accumulation has been selected because of their high quality performance for trust propagation in [8][61]. The trust attenuation is used to reduce the level of trust propagation as the number of links traversed increases from the trusted root set. Thus the effect of Trust attenuation is implemented using trust dampening. Thus each link has trust dampening factor β associated with it where $\beta < 1$. Thus this attenuation factor get multiplied for each link away from the root set of web pages and is used as multiplicative factor along with the trust value propagated from the parent node to the child node. Thus after trust propagation and attenuation, the trust based web graph G is created for a given cluster where each vertex in the graph is associated with the trust score.

In part B, HITS algorithm is applied on this trust based web graph for a given cluster where each vertex is associated with hub and authority score initialized using trust of the vertex. After the execution of HITS, each vertex in the web graph is categorized either as hub or authorities depending on the trust based hub and authority score. The vertex having high trust score for hub is more reliable hub than the reliable authority and vice versa.

Thus a given cluster is associated with trust based hubs and authorities as a result of completion of processing of Part A and Part B of this sub algorithm. The associated trust based hubs and authorities are further selected using the threshold value set for trust. This subalgorithm used for generating trust based Hubs and Authorities perform its task in the offline processing of the algorithm proposed for personalization of web search with trust based hubs and authorities.

The processing of Personalized Web Search based on Trusted Hubs and Authorities is divided into two phase: Phase I and Phase II. Phase I describes the offline preprocessing and Phase II describes the online processing of user search input queries.

Personalized Web Search based on Trusted Hubs and Authorities

Phase I

In Phase I, offline processing is performed in which all the tasks required for the execution of online processing of search input queries is implemented. The data set containing the input query and associated clicked URLs of the users on the web are preprocessed to get query session. The query

session keyword vector is generated from query session using Information Scent and content of the clicked URLs. These query session keyword vector are clustered using k-means algorithm. Trust is defined both at clicked URLs of the query sessions and at the cluster level. An subalgorithm for generating Trust based Hubs and Authorities using trust in HITS is applied on the web graph G generated for a given cluster of query sessions. Initially when the system generates no recommendations, the trust is undefined for all clusters therefore the information scent is used to select the relevant clicked URLs in a given cluster. The Graph is formed using selected clicked URLs in a given cluster in order to generate the High Scent Hubs and authorities using HITS. As the trust value of the clusters get defined with time, HITS is applied on web Graph generated using trusted web pages in a given cluster in order to identify trust based hubs and authorities of the cluster. The steps involved in offline processing are given below.

Algorithm:

Phase I:
Offline Preprocessing
<ol style="list-style-type: none"> 1. Data Set Collected on the Web is preprocessed to get the Query Sessions where each query session contains the user input query and associated clicked URLs. 2. For each clicked URLs in the query session, the Information Scent Metric is calculated using Eq. (1) which is the measure of the relevancy of the clicked URLs with respect to the information need of the user associated with the query session. 3. Query sessions keyword vector is generated from query sessions using Information Scent and content of Clicked URLs where content of clicked URLs is TF.IDF weighted vector see 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 $cluster_i_mean$. 6. The list L of clicked URLs for each cluster is created where $Information\ Scent(ClickedURL_i) \geq \rho(\text{threshold value})$. 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 $Trust(ClickedURL_i) = 0$ 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 $TrustDefined(i) = false$ and $Trust(i) = 0$

10. Invoke subalgorithm Trust in HITS(for generating trust based hubs and authorities) on the List L associated to each cluster of query sessions for computing Trust based Hubs and Authorities for each cluster.

SubAlgorithm:

Trust in HITS

Input: Clusters of query sessions and their associated List L of clicked URLs, Trust threshold value ϵ , Information Scent threshold value ρ .

Output: Clusterwise List L1 of Hubs and Authorities

For each cluster i do the following processing given in two parts: A and B.

Begin

Part A

1. If TrustDefined(i)=true then
 - a. Identify the clicked URLs in the list L associated with cluster i to form the root set R where Trust value $\text{Trust}(\text{ClickedURLi}) \geq \epsilon$.
 - b. The pages p in root set R is extended by following inlinks and outlinks on the web upto depth d to form the base set B.
 - c. The Web Graph $G=(V,E)$ based on base set B is created where V represents the set of web pages in the set B and E represents the link between web pages.
 - d. Trust is propagated in Web graph G using Logarithmic splitting from parent to child nodes and trust is accumulated on the child node using maximum share.
 - e. Trust value is attenuated using attenuation factor β as the number links traversed to reach the web page increases from the initial root set of web pages in R.
 - f. Thus each node p in the web graph G is associated with the Trust value which is used to initialize both hub y_p and x_p authority score of a node in the web graph G.

Else

- a. identify the clicked URLs in the list L associated with cluster i to form the root set R where Information Scent $\text{Scent}(\text{ClickedURLi}) \geq \rho$.
- b. The pages p in root set R is extended by following inlinks and outlinks on the web upto depth d to form the base set B.
- c. The Web Graph $G=(V,E)$ based on the base set B is created where V represents the set of web pages in the set B and E represents the link between web pages in the set B.
- d. Use the information scent of each page p in the root set R associated with the cluster i to initialize its authority weight

x_p and hub weight y_p of the corresponding vertex in the Web Graph G.

Part B

1. Apply HITS algorithm on the web graph G for a given cluster i obtained in Part A where the authority weight x_p and hub weight y_p for each node p in G are updated as follows till scores of each page p reach some fixed point.

$$x_p = \sum_{q \text{ such that } q \rightarrow p} x_q \quad (7)$$

$$y_p = \sum_{q \text{ such that } p \rightarrow q} y_q \quad (8)$$

2. Nodes p having high value of x_p than y_p will be viewed as good authority otherwise it is a good hub.
3. if TrustDefined(i)=true then
 Select the Hubs and Authorities in the web graph G associated with the selected cluster i whose trust based hubs and authorities score $\geq \epsilon$ and store it in list L1.
 Else
 Select the Hubs and Authorities in the web graph G associated with the selected cluster i whose Information Scent based hub and authorities score $\geq \rho$ and store it in list L1.
4. HITS output a short list L1 of the web pages with the high authority and hub score for a cluster i

End For

Phase II

During online processing, initially the user search input query is used to select the cluster which is most similar to the information need of the user. The selected cluster is used to recommend the associated Hubs and Authorities URLs in the List L1 and at the same time the recommended count of each recommended Hubs and Authorities is increased by one and Trustdefined status of the selected cluster becomes true if false. The clicked count of recommended HA URLs is increase by one for each click received by the user in the Personalized Search results. Thus trust metric is recomputed for each recommended HA URL in the selected cluster using recommended and clicked count. The trust metric of the recommended HA URLs in the selected cluster is further used to update the trust value of a given cluster. Once the trust value is defined and updated for clusters, then both the trust and cosine similarity measure are used in future to select the cluster for recommendations.

The steps involved in online processing are given below.

Phase II
Online Processing.
<ol style="list-style-type: none"> 1. The input query is used to find the most similar cluster. 2. For each cluster i the similarity is measured using the formulae $\text{MatchScore}_i(\text{input query}, \text{cluster}_i) = \frac{2(\text{sim}(\text{input query}, \text{cluster}_i\text{-mean}) * \text{Trust}(i))}{\text{Trust}(i) + \text{sim}(\text{cluster}_i\text{-mean}, \text{input query})}$ when $\text{TrustDefined}(i) = \text{True}$ $\text{sim}(\text{input query}, \text{cluster}_i\text{-mean})$ when $\text{TrustDefined}(i) = \text{False}$ 3. Identify the most matching cluster i 4. if $\text{TrustDefined}(i) = \text{true}$ then Identify the Hubs and Authorities in list $L1$ whose trust value $\geq \epsilon$ and ranked in decreasing order of their trust score. Else identify the Hubs and Authorities in list $L1$ whose Information Scent $\geq \rho$ and ranked in decreasing order of their Information Scent score. set $\text{TrustDefined}(i) = \text{true}$ endif 5. The Recommendedcount of the recommended HA URLs in list $L1$ are incremented by 1. 6. The user response to the recommended HA URLs is tracked and stores it in current user profile. 7. For each recommended HA URL clicked by the user, the ClickedCount of the corresponding recommended HA URLs in list $L1$ is incremented by 1. 8. The trust value of the selected cluster i and recommended HA URLs are updated as given below. $\text{Trust}(\text{HAURL}_i) = \{1 - \text{Distrust}(\text{HAURL}_i) \mid \text{where Recommendedcount}(\text{URL}_i) \neq 0\}$ $\text{Distrust}(\text{HAURL}_i) = \left\{ \frac{(\text{Recommendedcount}(\text{HAURL}_i) - \text{ClickedCount}(\text{HAURL}_i))}{\text{Recommendedcount}(\text{HAURL}_i)} \right\}$ Trust value of the selected cluster i is defined as follows $\text{Trust}(i) = \frac{ \text{CorrectSet}(i) }{ \text{RecSet}(i) }$ $\text{CorrectSet}(i) = \{ \text{HAURL}_i \mid \text{Trust}(\text{HAURL}_i) > \epsilon \mid \text{where Recommendedcount}(\text{HAURL}_i) \neq 0 \}$ $\text{RecSet}(i)$ is the total number of recommendations made using cluster i $\text{RecSet}(i) = \{ \text{HAURL}_s \mid \text{Recommendedcount}(\text{HAURL}_s) \neq 0 \}$ 9. If $\text{Trust}(i) = 0$ $\text{TrustDefined}(i) = \text{false}$

<ol style="list-style-type: none"> 10. If the user request for the next result page 10.1. 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 $\text{current_usersessionvector}_i$. 10.2. The similarity is measured for each i^{th} cluster using the formulae $\text{MatchScore}_i(\text{cluster}_i, \text{current_usersessionvector}_i) = \frac{2 * (\text{sim}(\text{current_user sessionvector}_i, \text{cluster}_i\text{-mean}) * \text{Trust}(i))}{\text{Trust}(i) + \text{sim}(\text{cluster}_i\text{-mean}, \text{current_usersessionvector}_i)}$ when $\text{TrustDefined}(i) = \text{true}$ $\text{sim}(\text{current_usersessionvector}_i, \text{cluster}_i\text{-mean})$ when $\text{TrustDefined}(i) = \text{false}$ 10.3. Goto step 3. 11. The updated trust scores of hubs and authorities in list $L1$ associated with each cluster is used to update the trust score of corresponding clicked URLs in the initial list L of URLs associated with the selected cluster. 12. Invoke the subalgorithm Trust in HITS offline at regular period of time in order to recompute the Hubs and Authorities associated with the clusters using the updated trust value of clusters and its List L of Clicked URLs. End
--

V. 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 capture the data set of clicked URLs of users using the search results of Google and performs the personalization of web search based on clustered query sessions. In order to generate the dataset of web user query sessions, 20 users volunteer to contribute to this experimental study and enter the input queries through a GUI based interface of the architecture. The search results of the input query 'hindi song' issued to the GUI interface of the architecture are retrieved from the web and displayed along with the check boxes as shown in Fig. 1 below. The user clicks on the retrieved search results are captured through the check boxes and are stored in the database in the form of query sessions for further processing.

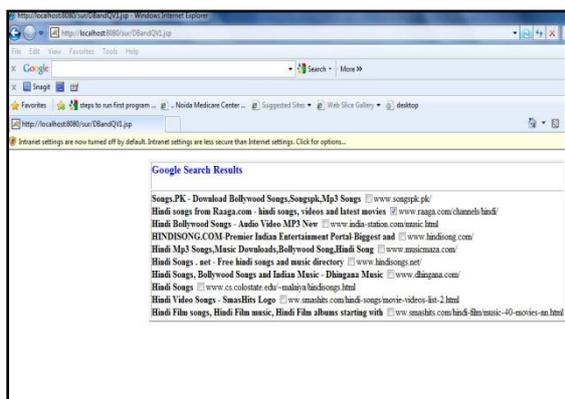


Fig.1. Screen Snapshot of architecture displaying Google Search results along with the checkboxes.

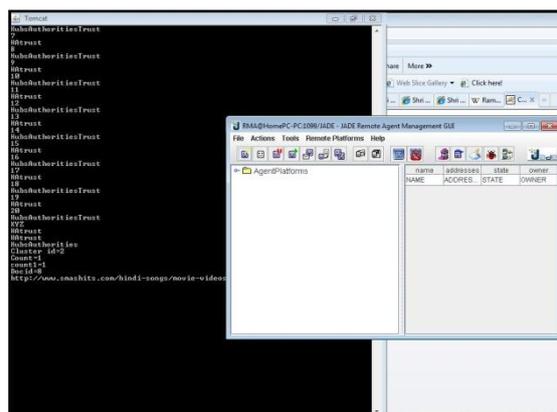


Fig.2. Screen Snapshot of execution of Hubs and Authorities Agent.

The experiment was performed on Pentium IV PC with 2 GB RAM under Windows XP. In the experimental set up for evaluating the performance of proposed approach of Personalized Web Search using Trust based Hubs and Authorities, the following values of the selected parameters shows the best performance. The threshold value of Trust ϵ was set to 0.5, threshold value of information scent ρ was set to 0.3, value of d (depth of crawling) was set to 4 and value of trust attenuation factor β was set to 0.5.

During offline processing, data set of user query sessions on the web is collected through the GUI of architecture, the Clustering Agent and Hub & Authorities Agent developed in JADE are executed to perform the processing involved in clustering and Hubs & authorities generation. The content 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 is executed to generate query session keyword vectors using Information Scent and tf.idf vector of the clicked URLs in the query sessions. The query session keyword vectors are clustered using k-means algorithm. It also performs the initialization of the trust of the clusters and the clicked URLs of the clusters.

The Hubs and Authorities Agent perform the processing associated with implementation of subalgorithm Trust in HITS. Trust in HITS is executed on each cluster of query sessions to generate the trust based hubs and authorities for each cluster. This Agent is invoked periodically at regular interval of time to work on the updated value of trust as the trust value of the clusters changes with time in response to the user clicks to the personalized web search results. Snap Shots of the execution of Hubs and Authorities Agent is shown in Fig. 2.

During Online processing, the input query is issued to GUI based interface designed to retrieve the search results using PWS with HA(with/without Trust) based on the same clustered query sessions dataset. Initially when the system has generated no recommendations, the trust associated with the cluster is undefined and PWS using Trust based Hubs and Authorities generates the Hubs and authorities recommendation using the Information Scent. But as the system generates the recommendations, the trust associated with the cluster gets defined and then the trustworthy cluster similar to the information need of the current user search query is selected. The resultant set of the trusted Hubs and Authorities associated with the selected cluster is recommended and displayed in the GUI Interface.

During evaluation of search results, users were divided into groups according to their expertise in the selected domains. The users were required to give the relevance score(0/1) to the search results. In PWS with HA (with trust), the recommended Hubs and Authorities are shown in decreasing order of their trust as shown in Fig. 3. Highly trusted Hubs and Authorities associated with the selected cluster are listed first. The user's clicks to the recommended Trust based hubs and authorities are tracked to capture the user's profile and dynamically update the trust associated with the recommended HAs and selected cluster.



Fig. 3. Screen Snapshot of Personalized Search Results using Trust Based Hubs and Authorities.

The Personalized results with Hubs and Authorities (without trust) are shown in Fig.4 given below. In Fig. 4. Recommended Hubs and Authorities are displayed in decreasing order of their Information scent.



Fig.4. Screen SnapShot of Personalized Search Results with Hubs and Authorities(without trust).

The performance of the PWS with HA(with trust) and PWS with HA(without Trust) proposed in [50] is compared using the average precision of personalized search results generated by each of the approach.

In order to evaluate the performance, the test queries were chosen in the domain Academics, Entertainment, Sports for covering wide range of queries on the web. The test queries were chosen randomly and there were 22 in Academics, 24 in Entertainment and 20 queries in Sports. During online searching, these test queries were issued in each of the selected domain to the GUI based interface of the architecture to retrieve the PWS with HA (with/without Trust). The precision of a given query using each of the PWS with HA(with/without trust) is computed by determining number of documents clicked by the user of the total retrieved documents. The average of precision of selected queries in a given domain is calculated for comparing the performance of PWS with HA(with/without trust) domainwise.

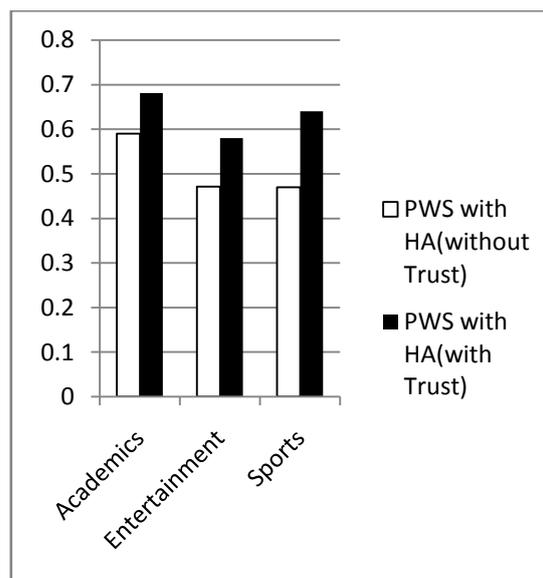


Fig.5. Compares the average precision of search results of PWS with Hubs and Authorities(HA) (with/without Trust) in Academics, Entertainment and Sports Domain.

The experimental results show the average precision of queries issued in each of the selected domain in Fig. 5. It is evident from the results that the average precision improves significantly for PWS with HA(with trust) in comparison to PWS with HA(without Trust). It is found when trust is used to generate the hubs and authorities in PWS with HA(with trust) the average precision of the personalized web search results is high and shows that trust based hubs and authorities were effective in satisfying the information need of the user. The fact that trust increases the reliability of HITS in identifying the good hubs and authorities is justified and provides the users high quality resource and content pages in a specific topic relevant to the information need of the user.

The obtained results were also analyzed using statistical paired t-test for average precision of PWS with HA(with trust) and PWS with HA (without Trust) with 65 degrees of freedom (d.f.) for a combined sample as well as for all three categories (Academics, Entertainment and Sports) with 21 d.f, 23 d.f and 19 d.f. The observed value of t for average precision was 22.6422 for a combined sample. Value of t for paired difference of average precision was 24.1174 for academics, 12.6984 for entertainment and 48.7983 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 PWS with HA (with Trust).

The results confirm that when both HA and trust are used in PWS the average precision is improved significantly in each of selected domain. PWS with trust based hubs and authorities retrieves higher number of relevant URLs in top ranked documents and increases their probability of being clicked by the users. Hence the increase in the ratio of relevant documents of the total documents retrieved is responsible for the improvement in the average precision in each of the selected domains. Thus PWS with trust based hubs and authorities personalizes the search more effectively than PWS with HA(without trust) with respect to the information need of the user.

VI. CONCLUSION

In this paper a novel approach is proposed for the personalization of web search using trust based hubs and authorities based on clustered query sessions. Experiment was conducted on the data set of query sessions captured in Academics, Entertainment and Sports domains to compare the performance of PWS with HA(with/without Trust). The performance is evaluated using the average precision of search results. The results verified statistically confirm the significant improvement in the precision in PWS when both HA and Trust is used in comparison to PWS with HA(without Trust). The results show that use of trust enhances the reliability of the HITS in identifying the good hubs and authorities. The average precision is improved as the number of relevant documents of the total retrieved documents is increased. Thus an increase in average precision results in effective Personalized Web search using trust based hubs and authorities catering to the information need of the user.

REFERENCES

- [1] A.Abdul-Rahman, S. Hailes. *A distributed trust model*, in: *Proceedings of the Workshop on New Security Paradigms*, ACM New York, U.S.A, 1997, pp. 48–60.
- [2] A.Abdul-Rahman, S. Hailes. *Supporting trust in virtual communities*. In *Proceedings of the 33th Hawaii International Conference on System Sciences*, IEEE computer Society, Washington, USA, 6, 2000, pp. 6007.
- [3] A.Jøsang. *Probabilistic Logic Under Uncertainty*, In *Proceedings of thirteen Australasian symposium on Theory of Computing*, 65, Australian Computer Society, Australia, 2007, pp. 101-110.
- [4] A.Kritikopoulos, and M. Sideri. *The Compass filter: Search engine result personalization using web communities*. *Lecture Notes Computer Science*, LNAI 3169, 2005, pp. 229-240.
- [5] A.Micarelli, and F. Sciarrone. *Anatomy and Empirical evaluation of an adaptive web-based information filtering system*. *Journal User Modeling and User-Adapted Interaction*, 14(2-3), 2004, pp. 159-200.
- [6] Ali Mohammad, Zareh Bidoki, and Nasser Yazdani. *DistanceRank: An Intelligent Ranking Algorithm for Web Pages*, *Information Processing and Management*, 44(2), 2007, pp. 877-892.
- [7] B. J. Rhodes. *Just-in-time information retrieval*. PhD. Thesis, MIT Media Laboratory, Cambridge, MA, 2000.
- [8] B. Wu, V. Goel and B. D. Davison. *Propagating trust and distrust to demote web spam*, In *Proceeding of Models of Trust for the Web (MTW)*, 2006.
- [9] C. Keser. *Experimental games for the design of reputation management systems*. *IBM Systems Journal*, 42(3), 2003, pp.498-506.
- [10] C.-N. Ziegler, and G. Lausen. *Paradigms for decentralized social filtering exploiting trust network structure*, *Lecture Notes in Computer Science*, Larnaca, Cyprus. Springer-Verlag, 3291, 2004, pp. 840–858.
- [11] C.-N. Ziegler, and J. Golbeck. *Investigating correlations of trust and interest Similarity*, *Journal Decision Support Systems*, 43(2), 2006, pp. 460-475.
- [12] D. Gambetta (Ed.). *Can We Trust Trust?* (Vol. 13). Oxford: University of Oxford, 2000.
- [13] David, Gibson, Jon, Kleinberg, and Prabhakar, Raghavan. *Inferring Web Communities from Link Topology*, In *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia: Links, Objects, Time and Space - Structure in Hypermedia Systems*, 1998, pp.225-234.
- [14] Direct Hit, Popularity-based search .1995. <http://www.directhit.com/>.
- [15] 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), 2002, pp. 35-59.
- [16] E H. Chi, 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*, ACM ,New York, 2001, pp.490-497.
- [17] Fabrizio Lamberti, Andrea Sanna, and Claudio Demartini. *A Relation-Based Page Rank Algorithm for Semantic Web Search*

- Engines, In IEEE Transaction of KDE, 21(1),2009, pp. 123-136.
- [18] F. Liu, C. Yu, and W. Meng. *Personalized Web search for improving retrieval effectiveness*, Journal IEEE Transactions on Knowledge and Data Engineering, 16(1), 2004, pp. 28 – 40.
- [19] H. Jiang, Yong-Xing Ge, and B. Han. *TIMERANK: A Method of Improving Ranking Scores by Visited Time*, In Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 2008.
- [20] H Kim, S. Lee, B. Lee, and S. Kang. *Building Concept Network-Based User Profile for Personalized Web Search*, In 9th International Conference on Computer and Information Science , 2010, pp. 567 – 572.
- [21] J.E. Pitkow, H. Schtze, T.A Cass, R. Cooley, D. Turnbull, A. Edmonds, E. Adar. and T.M. Breuel. *Personalized search. Communications of the ACM*, 45(9), 2002, pp. 50-55.
- [22] J. Golbeck. *Generating predictive movie recommendations from trust in social networks*, In Proceedings of the Fourth International Conference on Trust Management, Pisa, Italy, 2006, pp. 93-104.
- [23] J. Golbeck, J. Hendler. *Reputation network analysis for email filtering*, In Proceedings of the First Conference on Email and Anti-Spam, Mountain View, USA, 2004.
- [24] J. Golbeck, J. Hendler. *Filmtrust: Movie recommendations using trust in web-based social networks*, In Proceedings of the IEEE Consumer Communications and Networking Conference, 2006, pp.282-286.
- [25] 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, ACM New York, 2002, pp. 243-250.
- [26] J.M. Kleinberg. *Authoritative sources in a hyperlinked environment*. Journal of the ACM, 46(5), 1999, pp. 604-632.
- [27] J. O'Donovan, and B. Smyth. *Trust in recommender systems*, In IUI '05: Proceedings of the 10th International Conference on Intelligent User Interfaces, ACM New York, U.S.A, 2005, pp. 167–174.
- [28] Ko, Fujimura, , Takafumi Inoue, and Masayuki Sugisaki. *The EigenRumor Algorithm for Ranking Blogs*, In WWW 2005 2nd Annual Workshop on the Weblogging Ecosystem, 2005.
- [29] K.W.-T, Leung, W. Ng, and D.L. Lee. *Personalized Concept-Based Clustering of Search Engine Queries*, Journal IEEE Transactions on Knowledge and Data Engineering, 20(11), 2008, pp. 1505 – 1518.
- [30] Lian-Wang, Lee, Jung-Yi Jiang, ChunDer Wu, Shie-Jue Lee . *A Query-Dependent Ranking Approach for Search Engines'*, Second International Workshop on Computer Science and Engineering, Vol. 1, 2009, pp.259-263.
- [31] M. Kinateder, K. Rothermel. *Architecture and algorithms for a distributed reputation system*, In Proceedings of the First International Conference on Trust Management,1–16, Springer Verlag, 2003.
- [32] M. Montaner, B. L'opez, and J. Llu'is de la Rosa. *Opinion-based filtering through trust*, In Proceedings of the Sixth International Workshop on Cooperative Information Agents VI, Springer Verlag, London, UK, 2002, pp. 164–178.
- [33] M. Speretta,, and S. Gauch. *Personalized search based on user search histories*, In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, IEEE Computer Society Washington, DC, USA., 2005, pp. 622-628.
- [34] M. Vojnovic, J. Cruise, D. Gunawardena, and P. Marbach. *Ranking and Suggesting Popular Items*, In IEEE Transaction of KDE, 21(8), 2009, pp. 1133-1146.
- [35] N.L. Bhamidipati, K. Pal Sankar. *Comparing Scores Intended for Ranking*, In IEEE Transactions on Knowledge and Data Engineering,21(1), 2009, pp. 21-34.
- [36] O.E. Zamir, J.L. Korn, A.B. Fikes, and S.R. Lawrence. *Personalization of placed content ordering in search results*, United States Patent Application 20050240580, 2005. <http://www.freepatentsonline.com/y2005/0240580.html>.
- [37] P. Avesani, P. Massa, and R. Tiella. *A trust-enhanced recommender system application: Moleskiing*. In SAC '05: Proceedings of the 2005 ACM symposium on Applied computing, New York, 2005, pp.1589–1593.
- [38] P.A. Chirita, , J. Diederich, and W . NejdI. *Mailrank: Using ranking for spam detection*, In CIKM '05: Proceedings of the 14th ACM International Conference on Information and Knowledge Management, ACM New York, USA, 2005, pp. 373–380.
- [39] P. Bedi, H. Kaur. *Trust based personalized recommender system*, INFOCOMP Journal of Computer Science, 5(1) , 2006, pp. 19–26.
- [40] P. Bedi, S. Chawla. *High Scent Web Page Recommendation using Fuzzy Rough Set Attribute Reduction*. LNCS Transactions on

- Rough Sets XIV, J.F. Peters et al. (Eds.). 2011, LNCS 6600: 18-36.
- [41] P. Dasgupta. *Trust as a Commodity*. In D. Gambetta (Ed.), *Trust: Making and Breaking Cooperative Relations*. Oxford: Basil Blackwell, 2000.
- [42] P. O. Boykin, V. Roychowdhury. *Sorting e-mail friends from foes: Identifying networks of mutual friends helps filter out spam*. *Nature Science Updates*, 2004, 16.
- [43] 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, ACM New York, 1997, pp. 3-10.
- [44] 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.
- [45] R. Baeza-Yates, E. Davis. *Web page ranking using link attributes*. In Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, 2004, pp. 328-329.
- [46] R J. Wen, Y J. Nie, and J H. Zhang. *Query Clustering Using User Logs*. *Journal ACM Transactions on Information Systems*, 20(1), 2002, pp. 59-81.
- [47] R. Prajapati. *A Survey Paper on Hyperlink-Induced Topic Search Algorithms for Web Mining*, *International Journal of Engineering Research and Technology*, 1(2), 2012, pp. 13-20.
- [48] Sergey Brin , Larry Page. *The anatomy of a Large-scale Hypertextual Web Search Engine*, In Proceedings of the Seventh International World Wide Web Conference, 1998.
- [49] S. Chawla, P. Bedi. *Personalized Web Search using Information Scent*, In Proceedings CISSE'07 - 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), December 3-12, 2007.
- [50] S. Chawla, and P Bedi. *Finding Hubs and authorities using Information scent to improve the Information Retrieval precision* , In International Conference on Artificial Intelligence, WORLD COMP'08, 185-191, 2008.
- [51] S. Chawla. *Trust in Personalized Web Search based on Clustered Query Sessions*, *International Journal of Computer Applications* , 2012, 59(7), 36-44, Published by Foundation of Computer Science, New York, USA.
- [52] S. Chawla. *Personalized Web Search using ACO with Information Scent*, *International Journal of Knowledge and Web Intelligence*, 4(2/3), Inderscience Publishers, Geneva, Switzerland, 2013, pp. 238-259.
- [53] Shen Jie, Chen Chen , Zhang Hui, Sun Rong-Shuang, Zhu Yan and He Kun. *TagRank: A New Rank Algorithm for Webpage Based on Social Web*, In Proceedings of the International Conference on Computer Science and Information Technology, 2008.
- [54] S. Jin Kim, Sang Ho Lee. *An Improved Computation of the PageRank Algorithm*, In Proceedings of the European Conference on Information Retrieval (ECIR), 2002.
- [55] T. Dimitrakos. *A Service-Oriented Trust Management Framework.*, *International Conference on trust, reputation and security: theories and practice*, Springer –Verlag Berlin, 2002, pp. 53-72.
- [56] 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, 2006, pp.12.
- [57] Wenpu Xing, and Ali. Ghorbani. *Weighted PageRank Algorithm*, In Proceedings of the 2rd Annual Conference on Communication Networks & Services Research, 2004, pp. 305-314.
- [58] Xiang Lian, and Lei Chen. *Ranked Query Processing in Uncertain databases*, In IEEE KDD, 22(3), 2010, pp. 420-436.
- [59] Y. Zhao, and G. Karypis. *Comparison of agglomerative and partitional document clustering algorithms*, In SIAM Workshop on Clustering High-dimensional Data and its Applications, 2002a.
- [60] Y. Zhao, and G. Karypis. *Criterion functions for document clustering*, Technical report, University of Minnesota, Minneapolis, MN, 2002b.
- [61] Z. Gyöngyi, H.G. Molina, and J. Pedersen. *Combating Web Spam with TrustRank* , In Proceedings of the 30th VLDB Conference, Toronto, Canada, 2004, pp. 576-587.
- [62] Z. Zhu, J. Xu, X. Ren, Y. Tian, and L. Li. *Query Expansion Based on a Personalized Web Search Model*, In Third International Conference on Semantics, Knowledge and Grid, 2007, pp. 128 – 133.