IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 11, Issue 5 (May. - Jun. 2013), PP 56-69
www.iosrjournals.org

A Literature Survey on Ranking Tagged Web Documents in Social Bookmarking Systems

Nisar Muhammad, Shah Khusro, Saeed Mehfooz, Azhar Rauf
(Department of Computer Science, University of Peshawar, Pakistan)

ABSTRACT: *Social web applications like Facebook, YouTube, Delicious, Twitter and so many others have gained popularity among masses due to its versatility and potential of accommodating cultural perspectives in Social web paradigm. Social bookmarking systems facilitateusers to store, manageand share tagged web documents through folk classification system. These social toolsallow its users to associate free chosen keywords (tags) with documents for future considerations. Social tags reflect not only human cognition on contents the document contains inbut also used as index-terms in social searching. However search results associated with query-tags are randomly ordered either by popularity, interestingness or reverse chronological order with most recent bookmarks on top of search results, which limits the effectiveness of information searching in social bookmarking systems. A lot of research works have already been published to tickle the problem by exploiting different features of folksonomy structure. This survey provides a brief review of state-of-the-art, challenges and solutions towards recommending and ranking tagged web documents in Social Bookmarking Systems (SBS).*

Keywords*–Information Searching, Social bookmarking, Social Search, Ranking, bookmarks, Web 2.0*

I. INTRODUCTION

Due to the unbound storage nature of web, information searching and ranking has gained popularity in research communities. For this purpose search engines technologies and directorieswere implemented in 1990s, to overcome the issue of finding relevant information in search results. With the advent of Web 2.0 early in 2002, social indexing has greatly contributed in information management due to its informal organizational structure powered bythe online users of Web, where resources are associated with freely chosen terms instead of machine oriented controlled vocabularies.

Social bookmarking systems like Delicious and BibSonomy have large scale shared repository of public bookmarks enriched with social tags, provide tag-based information searching mechanism for facilitating users to search information they are interested in. However, search results returned by social searching systems are ordered either by popularity, interestingness or in reverse chronological order with recent bookmarks on top [1], [2]. The research problem is that bookmarked documents are not ranked according to their relevancy and importance to query-tag, which limits the effectiveness of searching information in social bookmarking systems. This survey reviews approaches proposed so far for re-ranking social search results against query-tags.

The survey is organized in different sections; section 2 provides a brief overview of information systems, Folksonomy and Social Bookmarking Systems. Section 3 is state of the art while Section 4 concludes the survey and proposes some questions and suggestions.

II. BACKGROUND

This section provides a brief overview about the background and historical development of information systems, Folksonomy and Social Bookmarking systems.

1.1 Social Information Systems

The basic objective of an Information System is to facilitate users with results having relevant documents on top of search results in decreasing order. Before the arrival of Internet and search engine's technologies most of research works were dedicated to centralized information retrieval systems, physically located at one centralized location. Information retrieval has been widely considered as a prominent research area since 1990s and after the technological development of search engines particularly. Since then a lot of research works have been done for example PageRank [3], HITS [4], SimRank [5], and SALSA [6] to improve IR systems. Similar approaches have also been adopted in social IR systems to enhance search results associated with query tags. Social bookmarking systems having large scale shared repository of public bookmarks, provide tag-based information searching mechanism for facilitating users to search information in public bookmarks they are interested in. However, search results returned by these systems are ordered either by popularity, interestingness or in reverse chronological order with recent bookmarks on top of search results. Information

searching and ranking requires quite sophisticated techniques in order to retrieve what is required.Section 3 reviews these approaches in the context of social web.

1.2 Folksonomy

The term Folksonomy was first coined by Thomas Vander Wall [7] and is the practice of collaboratively managing tagged web resources. Free chosen terminologies are used for annotations purpose instead of controlled vocabularies. According to A. Hotho [8] folksonomy is a quadruple F= (U, T, R, Y) where U, T and R represent set of users, tags and resources respectively, while $Y \subseteq U \times T \times R$ represent tag assignment, whereas the collection of all tag assignments is defined as folksonomy. The conceptual space is used for sharing, organizing and searching web resources in social web applications. Tags serve two purposes: locating web documents and provide qualitative data about contents. It represent contents of resources very well [9] [10], reflects human judgment on contents even without controlled vocabulary [11] and hence are considered valuable for information searching, indexing and ranking.Structure of folksonomy also called Formal Concept Analysis [12] has been widely discussed in different research articles as shown in Fig 1.1. S. Golden et al [13] have studied the structure of social tagging as well as its dynamic aspects. B. Lund and T. Hammond et al [14], [15] have investigated the structure of collaborative tagging system and architecture of participation. D. R. Millen et al [16] proposed an architectural design of social bookmarking tools for a large scale enterprise. R. Wetzker and N. Deka et al [17], [18] has reviewed main features, dynamics, patterns, tag spamming, and implications of tagging systems.

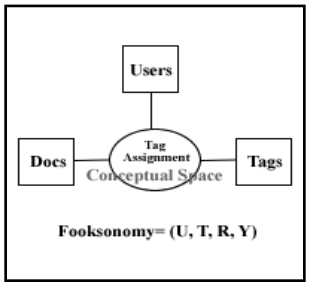

Figure 1: Structure of Folksonomy

1.3 Social Bookmarking Systems

Social Bookmarking Systems (SBS) is a Web 2.0 service using folksonomy to store web documents online, annotate with freely chosen terminologies, mark as private or public for sharing and future considerations. Some of these SBSs for example like Delicious, Connotea, Citeulike and BibSonomy have been reviewed by L. L. Barnes and F. Cevasco et al [19], [20].

Social bookmarking services are of great significance; beside informal organizational of knowledge it also creates communities of like-minded people for sharing interests as well as improve user's experience through expressing different perspectives and in-sighting others user's public resources [18][21]. Educational institutions, research communities also utilize online and offline learning, sharing [22][23] and Knowledge management by building up library services in the digital era through its democratic nature, allowing users to openly access and contribute [21][24][25].

III. SURVEY

Social Web has changed the way information is searched by incorporating human reasoning power with well-defined machine algorithms in social bookmarking systems. Potentials of social search have been studied in [26], [27]for enhancing web search. A lot of work has been published on recommending and ranking tagged web documents in order to improve web search by ranking relevant documents on top of search results against query-tag.wheremost of these techniques follows hybrid approaches but still categorized into six categories for reasonable organization and understanding: Personomy based techniques, Frequency and similarity based techniques, Structure-based techniques, Semanticsbased techniques, Cluster based techniques and Probability based techniques.

3.1 Personomy Based Technique

M. G. Noll et al [28] proposed the idea of ranking documents by exploiting similarity relationship between user's profile and document's profiles. Scalar-frequency based similarity is calculated among users, tags and documents by using equation (1).

$$\text{Similarity (u, d)} = p_u^T . p_d = \frac{\sum_l (tf_u(t_l).tf_d(t_l))}{\sqrt{\sum_l ((tf_u(t_l))^2}.\sqrt{\sum_l (tf_d(t_l))^2}} . p_d \qquad (1)$$

Where p_d is to dampen all the non-zero values to 1, the so called normalization factor. User profile is modeled by using bookmark collection as tag-document ($m \times n$) matrix M_d with m tags and n documents.

$$M_d = \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{m1} & \cdots & C_{mn} \end{bmatrix}, C_{ij} \in \{0,1\}$$

The value of C_{ij} is set to 1 if tag t_i is associated with document d_j or otherwise 0. Each user profile is thus formed by equation (2).

$$P_u = M_d.\omega_d = \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{m1} & \cdots & C_{mn} \end{bmatrix}.\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} C_i^* \\ \vdots \\ C_m^* \end{pmatrix}, \; C_i^* \in N_0 \qquad (2)$$

The $\omega_d = I^T$ represent [1…1] and factor C_i^* represents the total count for tag t_i in user's bookmarking collection for users. Similarly each document profile is $P_d = M_u.\omega_u$ Where M_u is tag-user matrix of ($m \times u$).

The same efforts has been made by S. Xu et al [29] where the normalized $tf * idf$ based angular distance between user's and document's profiles is computed by equation (3):

$$\text{Similarity} = COS_{tf-idf}(u_m, d_n) = \frac{\sum_l (tf_{u_m}(t_l).iuf(t_l).tf_{d_n}(t_l).idf(t_l))}{\sqrt{\sum_l ((tf_{u_m}(t_l).iuf(t_l))^2}.\sqrt{\sum_l (tf_{d_n}(t_l).idf(t_l))^2}} \qquad (3)$$

This combines the term matching between query and web page with topic matching between users and web pages. Where $tf_{u_m}(r_l), tf_{d_m}(t_l)$ represents user based frequency and document frequency respectively. The parameters $iuf(t_l), idf(t_l)$shows user based and document based inverse documents frequency. Whereas $|u_m|, |d_n|$ are user's and document's magnitudes.

D. Vallet et al [30] modified the above scheme by eliminating the user and document length normalization factors using equation (4).

$$tf - idf(u_m, d_n) = \sum_l (tf_{u_m}(t_l).iuf(t_l).tf_{d_n}(t_l).idf(t_l)) \qquad (4)$$

The work is different from that of [28] in the sense, that it uses iuf and idf in combination for tag distribution globally for users and documents. These personalized approaches has been summarized and compared by I. Cantador et al [31].

Y. Cai et al [32] defined user profile and document profile as:

$$\overrightarrow{U_i} = (t_{i,1} : v_{i,1}, t_{i,2} : v_{i,2} : \ldots : t_{i,n} : v_{i,n})$$
$$\overrightarrow{D_C} = (t_{c,1} : \omega_{c,1}, t_{c,2} : \omega_{c,2} : \ldots : t_{c,n} : \omega_{c,n})$$

Where $v_{i,x} = \frac{N_{i,x}}{N_1}$ is the ratio of the number of times user i used tag x for resource N with total number of resources tagged by user i, the NTF for tag x used by user i. Whereas $\omega_{c,x} = \frac{M_{c,x}}{M_c}$ is the count of users who annotate resource c with tag x divided by the total number of user who ever annotate resource c with any tag. The personalized ranking function in equation (5) is based on user and document profiles as.

$$RScore(\overrightarrow{q_i}, \overrightarrow{U_i}, \overrightarrow{D_C}) = \frac{\gamma(\overrightarrow{q_i}, \overrightarrow{R_C}) + \theta(\overrightarrow{U_i}, \overrightarrow{R_c})}{2} \qquad (5)$$

The first part $(\overrightarrow{q_i}, \overrightarrow{D_C}) = \frac{\sum \omega_{c,x}}{m} . \left(\frac{k}{m}\right)^{\alpha}, t_{c,x} \in \vec{q}_i$ Where k the total number of terms in query satisfied by resource profile and m represents total number of terms in query. □is the constant used to adjust the effect of relevant tags in a resource profile for query. And $\theta(\overrightarrow{U_i}, \overrightarrow{D_j}) = \frac{\sum l_x v_{i,x}}{m}$, m is the total number of terms in the query, and

$$l_x = \begin{cases} \omega_{c,x} + (1 - v_{i,x})(1 - \omega_{c,x}) & 1 > \omega_{c,x} > 0, \; v_{i,x} > 0 \\ 1 & \omega_{c,x} = 1, v_{i,x} > 0 \\ 0 & \omega_{c,x} = 1, v_{i,x} > 0 \end{cases}$$

P. Wu et al [33] proposed personalized recommendation by exploiting $tf * idf$ weightage as a variable value for diffusion based algorithm for personalized recommendation and ranking for document j using equation (6).

$$r_{k,j} = \sum_{l \in \Gamma(i_j)} (p_l * w_{l,j}) \qquad (6)$$

The first factor represents $tf * idf$ as $w_{k,j} = w_{k,j}^{(T)} + \in$, and the second factor is preference factor having the weight of the edge between $u_k \& i_j$. The weight of item i_j with respect to user u_k is defined by equation (7):

$$w_{k,j} = w_{k,j}^{(T)} + \in = \sum_{t \in \Gamma(k,j)} (freq_{kt} * \log \frac{|U|}{|\{u : t \in \Gamma(u)\}|}) + \in \qquad (7)$$

$\Gamma(k, j)$, is the personal tag space of user $u_k = (k = 1,2,3, \ldots, n)$, $freq_{kt}$ is the frequency of tag t used by u_k. $|U|$ is the total number of users in system, $\Gamma(u)$ is the set of tags used by user k and $\{u: t \in \Gamma(u)\}$ count the number of users who used tag t.The diffusion process first defined in equation (8), distributes values calculated for each item, averagely to the users who have collected; each user u_l receives a value as:

$$pl = \sum_{j \in \Gamma(uk)} \left(\frac{w_{k,j}}{d(i_j)} \right) = \sum_{j \in \Gamma(uk)} \left(\frac{\text{weight of document } j}{\text{count of users who have collected the document } j} \right) \quad (8)$$

$\Gamma(uk)$ is the set of items that have been collected by uk, $d(i_j)$ is the degree of item i_j in the user-item bipartite network. In step 2, the value of each user is redistributed among its item's collection according to $w_{k,j} = w_{k,j}^{(T)} + \epsilon$.

H. N. Kim et al. [34] proposed Folksonomy-boosted ranking (FBR) which follows collaborative filtering mechanism for personalized ranking, the hybrid approach is given by equation (9):

$$FBR_u(i, q) = \sum_{t \in q} (p_{u,t} * w_{t,i}) \quad (9)$$

The latent tag preference P defines the dot product of user-tag matrix A with tag-tag similarity matrix E using equation (10).

$$P = \tilde{A} \cdot E^k \quad \text{OR} \quad P_{u,t} = \tilde{a}_u^{|T|} \cdot e_t \text{ OR } P_{u,t} = \sum_{J=1}^{|T|} \tilde{a}_{u,j} \times e_{j,t} \quad (10)$$

Where $\tilde{A} = \tilde{a}_{u,t} = a_{j,t} \Big/ \sqrt{\sum_{J=1}^{|U|} (a_{j,t})^2}$ is the normalized matrix of A, The second part E^k shows k most similar tags $e_{j,t}$ computed as $e_{x,y} = \cos(t_x . t_y) = t_x \cdot t_y \Big/ ||t_x|| \, ||t_y||$

The latent tag annotation model is given by equation (11).

$$W = \tilde{N} . H^k \text{ OR } P_{u,}i = \tilde{n}_t^T . h_t \quad \text{OR } P_{u,}i = \sum_{J=1}^{|t|} \tilde{n}_{u,j} \times h_{j,i} \quad (11)$$

N is tag-item matrix which values n_{ti} represent number of users who have annotated item i with tag t, computed as $\tilde{n}_{t,i} = n_{t,i} \Big/ \sqrt{\sum_{J=1}^{|t|} (n_{i,j})^2}$, H is item × item matrix and H^k contains K most similar items. The values H^k for document x and y is calculated, $h_{x,y} = \cos(i_x . i_y) = \frac{i_x \cdot i_y}{||i_x|| . ||i_y||}$.

3.2 Tag Frequency and Similarity Based Techniques

Theseranking techniques which follows similarity and frequency based approaches are reviewed in this section. Mostly Vector Space Model is used to calculate similarity measures between web pages and query.

3.2.1 CoolRanking

H. S. Khalifa et al [35] proposed CoolRank to rank tagged documents by exploiting the relationship among tags, resources, and users. $CoolRank$ Algorithm is based on two assumptions, resource popularity *P(R)* and tag subjectivity S(R) by using equation (1).

$$\text{CoolRank} = S(R) + P(R) = \frac{\sum ft(T) \in \text{Tags}}{U(R)} + \log(\text{No: of people who bookmarked Resourece R}) \quad (1)$$

ft(T) is the occurrences of tag query-tag T and U(R)$is\ total\ bookmarks\ of\ user\ U$.

3.2.2 Social Ranking

V. Zanardi et al [36] proposed Social Ranking by exploiting the cosine similarity relationships among tags and users. Social rank for a document is the sum-total of tag similarity and user's similarity using equation (1):

$$R(d) = \sum_{u_i} ((SimTags) * sim(\bar{u}, u_j)) + 1 \text{ OR}$$

$$R(d) = \sum_{u_i} \left(\sum_{\{t_x | u_i \text{ tagged } p \text{ with } t_x\}, \{t_j \in q^*\}} sim(t_x . t_j) \right) * sim(\bar{u}, u_j) + 1 \quad (1)$$

The proposed technique first expand query-tag by considering tags that are similar to query q for which $0 < sim(t_i . t_j) < 1$ where $t_i \in q$ and $t_j \in q'$, q´ is the set of similar tags. The social rank score for a document $R(d)$ is the combination of relevance of tags associated with publication with respect to tags in the extended query set q^* and the similarity of the users with respect to query user; count the number of similar users.

3.2.3 Normalized Match Tag Count (NMTC)

W. Choochaiwattana at al [37] proposed MTC and NMTC expressed by the equations (1) and (2). MTC (Match Tag Count) calculates number of users who used tags matches with query terms string. That is $M_{rua=}(R_x, U_y, A_z)$ equal to 1 when user U_y uses tag A_z to annotate resource Rx or otherwise as stated. The set $q= \{q_1,q_2,q_3, q_4 \dots q_n\}$ represents n query terms and $a(r_x)=\{a_1,a_2,a_3,a_4,\dots,a_n\}$ is the annotations set of web resource x. $NMTC_x$ (Normalized Match Tag Count) is the normalized form of MTC_x which count all matched tags for a resource.

$MTC_x = \sum_{y=0}^{N_u} \sum_{z=0}^{N_a} M_{rua}(r_x, u_y, a_z)$ *if* $a_z \in q$ . . . . . (1)

$NMTC_x = \frac{\sum_{y=0}^{N_u} \sum_{z=0}^{N_a} M_{rua}(r_x,u_y,a_z) \text{ if } a_z \in q}{\sum_{y=0}^{N_u} \sum_{z=0}^{N_a} M_{rua}(r_x,u_y,a_z)}$ . . . . (2)

3.2.4 Social Ranking Based on Reputation

E. M. Daly et al [38] exploited the Wisdom of the Crowds so called reputation, which combinations the number of bookmarkers, reputation of the bookmarkers and time dynamics of documents in order to rank web documents. User reputation is the number of users (consumers) consuming the content of a user (contributor). The factor R_{reward} depends upon the consuming rate of other users by equation (1).

$User\ Reputation = R_{new} = R_{old} + (1 - R_{old}) \times R_{reward}$ . . . . (1)

The document reputation is simply the number of users that add documents to their collection using equation (2) and time dynamic by equation (3).

$Document\ Reputation = R_{new} = R_{old} + (1 - R_{old}) \times R_{reward}$ . . . . (2)

$Time\ Dynamics = R_{new} = R_{old} \times \gamma^k$ . . . . . (3)

The third component is the time dynamic, where γ shows time decay coefficient and k is the time unit since the reputation value for an item last updated. The Reputation ranking is given in equation (4).

$Reputation\ Raning = R_{new} = R_{old} \times R_{bookmarker} \times \beta$ . . . . . (4)

3.2.5 Tag-Similarity Based Ranking

F. Durao et al [39] proposed tag-based system by suggesting similar web pages based on the similarity of their tags and a reordering method of the original recommended ranking. Three arguments are combined to evaluate the personalized recommendation ranking; tag popularity, tag representativeness and affinity of tags-users. The document score is given by equation (1).

$Ds = \sum_{i=1}^{n} weight(Tag_i) * \sum_{i=1}^{n} representativness(Tag_i)$ (1)

where n is the total number o f existing tags in the repository

The affinity between user and tag is calculated with equation (2):

$Affinity_{(u,t)} = \frac{card\{r \in Documents \mid (u,t,r) \in R, R \subseteq U \times T \times D\}}{card\{t \in T\ (t,u) \in R_u, R_u \subseteq U \times T\}}$, (2)

where t is a particular tag, u particular user, U is the set of Users and D is the set of resurces and T is the set of tags.

By combining the above three parameters we have the following equation:

$Similarity_{(D_i,D_j)}$

$= \left[Ds_{D_i} + Ds_{D_j} * cosine\_similarity\left(T_{D_i}, T_{D_i}\right)\right]$

$* Affinity_{(u,t)}$, *Where Ds is the document score and T is the set of tags of a particular document.*

3.2.6 Tensor Based Recommendation and Ranking

R. Wetzker et al [40] proposed user-center tag model (UCTM) which maps personomies into folksonomies. Ternary relation is utilized among items, users and tags called the folksonomy tensor $Y = Y_{itu} = \subseteq I \times T \times U$ with$Y_{itu} = 1$, *if* $(i,t,u) \in Y$ *and* 0, *Otherwise.* Ranking of the proposed technique took place in two steps: first query tag is translated to the global vocabulary and then items with highest weight are recommended to users. Translation of a single tag t is the previous co-occurrences with other tags represented by vector$\tau_{T\tilde{t}u}$ within the translation tensor$\tau_{T\tilde{T}u}$. Tags from user's personomy may be translated to folksonomy vocabulary by simple vector multiplication using equation (1).

$t(\tilde{t}, u) = \frac{\tau_{T\tilde{T}u} \bar{\times}_T \tilde{t}}{|\tau_{T\tilde{T}u} \bar{\times}_{\tilde{T}} \tilde{t}|}$ (1)

In step second items which are associated with these community tags are ranked by calculating weight vector by using the following equation (2).

$\hat{\imath}(\tilde{t}, u) = \frac{\tau_{T\tilde{t}u} \times_T A^* IT}{|\tau_{T\tilde{t}u} \times_T A^* IT|}$ (2)

Where $A^*IT$is the tag-normalized stochastic version of the matrix AIT.

3.2.7 Neighbor Weight Collaborative Filtering (NwCF)

D. Parra-Santander [41] proposed the idea of Collaborative Filtering in Social Tagging Systems by exploiting the collaborative filtering recommender system [42]. The proposed technique works in two steps: first users with similar interests are filtered by using equation (1).

$$userSime(u,n) = \frac{\sum_{i \subseteq CR_{u,n}} (r_{ui} - \bar{r}_u)\,(r_{ni} - \bar{r}_n)}{\sqrt{\sum_{i \subseteq CR_{u,n}} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{i \subseteq CR_{u,n}} (r_{ni} - \bar{r}_n)^2}} \ldots . (1)$$

The above equation shows user's similarity in terms of Pearson Correlation coefficient between target user u and neighbor n whereas $CR_{u,n}$ shows set of all correlated items between u and n, and r_{ni} represents neighbor's n rating for item $i \in CR_{u,n}$. Top k neighbors are considered and items of those neighbors are recommended to the target user.

$$pred(u,i) = \bar{r}_u + \frac{\sum_{i \subseteq neig\ hbors\ (u)} userSime\ (u,n).(r_{ni} - \bar{r}_n)}{\sum_{i \subseteq neig\ hbors\ (u)} userSime\ (u,n)} \ldots . . (2)$$

Enhancement of the equation (2) is supplemented by $nbr(i)$ as described below in equation (3).

$$pred'(u,i) = \log_{10}(1 + nbr(i)).pred(u,i) \ldots . (3)$$

The neighbors of top k users (neighbors) are calculated by using equation (1). Here $nbr(i)$ calculates the number of raters in the overall calculation of publications.

3.2.8 Linear Weighted Hybrid Approaches

J. Germmell et al [43] proposed different hybrid approaches based on users and items based collaborative filtering, tag model similarity, structure and popularity.

a) Linear Hybrid Recommendations using equation (1)

$$\emptyset(u,q,r) = \sum_{i=1}^{k} \alpha_i \emptyset_i(u,q,r) \qquad (1)$$

The component which exploits the two-dimensional properly of URT model include RT, RU, TU, TR, UR, UT, for example $RT(r,t) = \sum_{\forall u \in U} URT(u,r,t) = The\ number\ of\ users\ who\ have\ annotatedrwitht.$

b) Collaborative Filtering:

User based collaboration is based on KKN ($KKN_{ru}\ and\ KKN_{rt}$) collaborative filtering algorithms. User-based tag-specific approach is defined by equation (2):

$$\emptyset(u,\{t\},r) = \sum_{v \in N_u^t} \sigma(u,v)\chi(v,t,r)\ where\ \theta(v,t,r) = 1\ if\ v\ annotate\ r\ with\ t\ or\ 0\ otherwise. \qquad (2)$$

$\emptyset(v,t,r)$ Calculated by computing the similarity measure between users (target user u and neighbor v), where user v must have at least label resource r with tag t. In case of item-based collaborative filtering recommendation systems we have equation (2):

$$\emptyset(u,\{t\},r) = \sum_{s \in N_r^t} \sigma(r,s)\ \theta(u,t,s) \qquad (3)$$

Similarities between resources are computed between the given resource r and neighbor's resources $s \in N_r$. A resource is considered only if tagged with tag t for finding k neighbors for a resource.

c) Tag-based Similarity Model:

Tag based similarity model defined by the following equation (4).

$$\omega(u,\emptyset,r) = \frac{\sum_{t \in T} RT(r,t) \times UT(u,t)}{Magnitude\ \ of\ RT(r,t).Magnitude\ \ of\ RT(u,t)} \qquad (4)$$

d) Popularity Model: Tag specific popularity model is defined by equation (5).

$$\omega(u,\{t\},r) = \sum_{v \in U} \chi(v,t,r) \qquad (5)$$

The value of $\chi(v,t,r)$is 1 only if user v tag resource r with t and zero otherwise.

3.3 Structured Based Approaches

Different mathematical models like Matrices, Functions, Vectors, Probability and Graph have been considered in computing, to make the transformation of various research concepts into real world practices possible. The graph technique has been adopted in PageRank, HITS and SALSA as far as search engines technology is concerned. In the same way graph models are utilized to associates different objects of social bookmarking systems for the sake of organization, sharing and ranking e-resources. Some of the contributions that exploit graph structure of folksonomy are analyzed in this section.

3.3.1 FolkRank

FolkRank is the adapted $RageRank\ [\vec{\omega} = Ad\vec{\omega} + (1-d)\vec{p}]$ algorithm for folksonomy based information ranking proposed by A. Hotho and his team [1]. The folksonomy $F = (U,T,R,Y)$ structure is converted into an undirected tri-partite graph as $G = (V,E)$ as shown in figure 2. The edge $W_{(t,r)} = \{u \in U\}$ shows number of users who annotated resource r with tag t, the relation $W_{(t,u)} = \{ r \in R \}$ represents the number

of resources tagged with term t by user u and $W_{(u,r)}$= { $t \in T$ } is the number of tags assigned to resource r by user u. All these weightages mutually reinforce each other by spreading their weights using equation (1).

$\vec{\omega} = Ad\vec{\omega} + (1-d)\vec{p}$ (1)

Where A is the adjacency matrix representing weight-of-edges between nodes and is a model of the folksonomy graph, all the in-links contributes an input value in calculating the overall $FolkRank$ of a node. p is the preferences vector and d is dumping factor. $FolkRank$ is calculated by differentiating two computation of $PageRank$, one with and one without a preference vector as $\vec{\omega} = \vec{\omega}_1 - \vec{\omega}_0$ where $\vec{\omega}_0$ is the where $d = 1$ and $\vec{\omega}_1$ is with$d < 1$. The final differential vector contains the F$olkRank$ of each node; tag, resource and user.

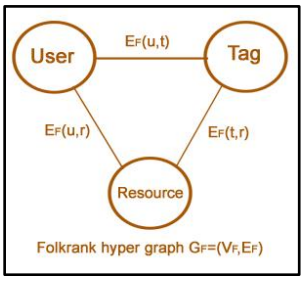

Figure2: Tri-Partite graph of tags, users and URLs

3.3.2 SocialSimRank (SSR) and SocialPageRank (SPR)

S. Bao et al [44] proposed SSR and SPR Algorithm for measuring the popularity of web resources from user's perspective. SSR is computed based on set of tags associated with a page and query terms by calculating SimRank [5] similarity measure between query and tags as using equation (1).

$Sim_{SSR}(q,p) = \sum_{i=1}^{n}\sum_{j=1}^{m} S_A(q_i, q_j)$ (1)

The queryq_i having many terms and q_j is the annotations set associated with web page p..The SPR calculates popularity of web pages in the context of social bookmarking by executing the following steps from (a) to (f):

a) $U_i = A_{PU}^T . P_i$
b) $A_i = A_{UA}^T . U_i$
c) $P_i^{'} = A_{AP}^T . A_i$
d) $A_i^{'} = A_{AP} . P_i^{'}$
e) $U_i^{'} = A_{UA} . A_i^{'}$
f) $P_{i+1} = A_{PU} . U_i^{'}$ Until P_i converge which is the SPR score for resources.

The popularity vectors are P_i , U_i and A_i represent pages, users and annotations in the ith iteration. User popularity can be derived from the pages they annotate at equation (a), the annotation popularity can be derived from the popularity of users in equation (b), the popularity of web pages can be derived from annotations in equation (c), and Web pages to annotations at equation (d), annotations to users at equation (e) and users to web pages at equation (f).

3.3.3 Group Sensitive Ranking

Group-Sensitive ranking algorithm is based on $GroupMe$ Folksonomy [45]. A $GroupMe$ folksonomy is defined as a 5-tuple $F = \{U, T, R, G, Y\}$ where U, T, R, Y represent users, tags, resources and tag-assignment respectively, while G = { g_1,g_2,g_3,…, g_n} is the set of groups like g_1 or g_2 ∈G whereas a group is a set of resources. The factor Y represents the tag assignment in the context of $GroupMe$ with $Y \subseteq (U \times T \times R \times U)$. The $GroupMe$ users can create groups of topic specific resources which are related to each other. During the survey it was observed that 50% resources do not have even a single tag. Therefore it is hard to found information by folksonomy based searching. The $GroupMe$ approach allows users the facility of free for all tagging approach which enables users to annotate not only their own resources but resources of others as well by annotating groups having many resources [46].

3.3.3.1 GRank

The technique proposed by F. Abel et al [47] is a group sensitive ranking method operates on $GroupMe$ Folksonomy environment where resources are grouped on the basis of similarity and users interests. For query tag q_t, group $g \in G$ the ranking vector returns $\vec{w}_{Rq}(r)$ weights as $GRank$ for resource $r \in R_q$ by the following four factors:

a. $\vec{w}_{Rq}(r) = w(tq, r) \ .d_a$ [$weights\ of\ directely\ annotations$]
b. For each group g$g \in G \cap d_b$ compute:

$\vec{w}_{Rq}(r)+= w(tq,g)\ .d_b\ [weight\ gained\ from\ goupe]$

c. For each $r^{'} \in R_a$ where $r^{'}$ is contained in the same group as that of r and $r \neq r^{'}$ do$\vec{w}_{Rq}(r)+= w(tq,r^{'})\ .d_c\ [weight\ from\ neighbor\ resources]$

d. If $r \in G$ then: for each $r^{'} \in R_a$ where $r^{'}$ is contained in r do: $\vec{w}_{Rq}(r)+= w(tq,r^{'})\ .d_d\ [weight\ of\ reourece\ in\ goupe]$

3.3.3.2 GFolkRank: Group Sensitive FolkRank

GFolkrank [48] is the adopted FolkRank which is a context sensitive ranking algorithm operates on graph $GG = (VG, EG)$ which models $F = (U, T, R, Y, G)$. First F is transformed to GG where each node contributes to every other node recursively. When a user u adds a resource r to a group g, the tag assignment Y is formulated as (u, tg, r), where t_g belongs to TG. These tags are called artificial tags which are assigned to all resources containing in a group g. In this way the vertices of hyper graph is increased by TG having a total of $VF + TG$vertices.

3.3.3.3 Social HITS: Social Hyperlink Induced Topic Search

F. Abel [49] proposed Social HITS operates in group sensitive folksonomies. In the GroupMe context the hub and authorities which are not constraints to entity types: for example user's authority may be supposed as the annotations to high quality resources, similarity tag authority may the contributed by high quality users etc. The algorithm executes in:

a) Folksonomy model$F$, t the query, St the searching strategy, Sg the graph construction strategy and k the number of iterations.
b) Search base set (F_t) tag-assignments relevant to t.
c) Construct graph (G_D) from the base set (F_t) by using graph construction strategySg.
d) Iterate $G_D$graph k times for calculating Hub and Authorities of resources in F_t.

3.3.4 ExportVoteRank and RecommendationPageRank

C. H. Lo et al [50] proposed $ExportVoteRank$ (EVR) and $RecommnedationPageRank$ (RPR) for ranking resources in social tagging environments. The EVR for a web resource considers the importance of a resource by taking into account authorities of users who annotate it using equation (1).

$$evr(r) \leftarrow evr(r) + \frac{pr\,(u)}{count\ (u)} \qquad (1)$$

$\forall u \in U, \forall (u,r) \in UR, UR = \{(u,r) | \exists t \in T\ such\ that\ (u,t,r) \in UTR\}$, $Pr(u)$ is Rank score of user u and initially set to 1 and$Count(u)$ to zero as number of bookmarks by u.

RecommendationPageRanking is based on association rule mining and implemented as a directed graph RecGraph $G_t = (V_t, E_t)$ among users, tags and resources for$(V_t, E_t) \in UR_t, the\ set\ of\ resources\ r \in R\ related\ to\ query - tag\ t$. The$pr_t(r_i)$as PageRank for $r_i \in V_t$is computed in equation (2):

$pr_t(r_i) \leftarrow \frac{1-d}{\|V_t\|}$+ (d) . $\sum_{r_j \in E_t(r_i)} \omega_t(r_j, r_i).pr_t(r_j)$ With

$$rpr(r) \leftarrow rpr(r) + pr_t(r) \quad \forall t \in T, \forall r \in V_t \qquad (2)$$

3.3.5 S-BITS: Social Bookmarking Induced Topic Search

The idea of HITS (Hypertext Induced Topic Search) is followed in the S-BITS [51] approach toward ranking search result in social bookmarking system. Basic assumptions which drive the proposed technique are:

a) A resource tagged by many good users is a good resource and
b) A user annotated many good pages is a good user

The folksonomy model is exploited to exhibit the phenomenon of hub and authority on bookmarked documents. The HIT algorithm works as follow:

a) Input tag-query terms k to a text-based IR system and obtain the root-set R by using some similarity measures, Web pages that are collected, as Root Set R, are related to the query k. Additionally the following is formatted.

a) All the users associated with the root set R are taken in account to make set U of users, where each user $u \in U$ have at least one annotation to a page in R.
b) All the tags associated with R are considered as bookmarked-tag set BT for$R$.

b) The root-set R is expanded by using association rules to obtain a base set B as $B = R \cup \{p_j | u_i \xrightarrow{BT_{ij}} p_j \wedge u_i \in U \wedge FT \subseteq BT_{ij} \wedge FT \in T^{'}\}$. Where $FT \in T^{'}$ is the frequent tag set and each tag set $BT_{ij} \in T^{'}$ is taken as a transaction with association rule$u_i \xrightarrow{BT_{ij}} p_j$.

c) The authorities and hub score are calculated as per the nested graph that is shaped by web pages in B and users in U and annotation as edges E.
d) Report top-ranking authorities and hubs.

3.4 Semantic Based Approaches

Semantic web so called Web 3.0 defined as web with a meaning [W3Schools], which is still not fully implemented and is a web of things described syntactically in a way that computer can understand [52]. It refers to different methods and technologies likes Resource Description Framework [53], data interchanging formats like RDF/XML, Triple and notation like RDF Schema and Web Ontology Language (OWL) [54] that makes the Web more intelligent by providing formal description (via Description logic and other knowledge representation techniques) of concepts and relationships within a given knowledge domain.

3.4.1 GroupMe With Semantics Approach

F. Abel et al [55] proposed ranking approach based on social semantic web, considering not only tag frequency but also contextual information in the form of groups been enriched with RDF semantics. The operations executes on group folksonomy in the following sequence.

a) All resources that are annotated with $t_i \in q= \{t_1, t_1, .. t_n\}$ are retrieved and weight of each resource is calculated as equation (1):

$$\text{resourceWeight}(t,r)=\sum_{t\in q} \text{resourcesWeight}(t,r) = \frac{\text{number of users who tagged resource r with t}}{\text{number of users who tagged resource r}} \quad (1)$$

b) All the groups are consider which are tagged with $t_i \in q= \{t_1, t_1, .. t_n\}$ where group weights are calculated for each page in a group using equation (2).

$$\text{groupWeight}(t,g) = \frac{\text{number of resources in g that are tagged with t}}{\text{number of resouces in g}} \quad (2)$$

c) Context weight is calculated for documents based on their appearances in groups based on Equation (a) and (2) using equation (3):

$$\text{ContextWeight}(q, r, g) = \sum_{t\in q} \text{resourcesWeight}(t,r) \,.\, \text{groupWeight}(t,g) \quad (3)$$

Group-Me is a semantic application which extracts concepts like groups and relations of tags in and out of groups from IR systems to transform them into RDF descriptions with the help of DC, FOAF and Group-Me and provide semantically reach description for the group folksonomy.

e) Resources retrieved are ranked according to their weightage

3.4.2 Semantically Relevant Resource Retrieval and Ranking (SR3)

P. Bedi et al [56], [57] proposed SR3 that exploits advantages of social bookmarking services and Semantic Ontologies. Query is expended by using domain ontologies, where query-tag Qo becomes $Q = \{Q_0, Q_p, Q_c\}$. SR3 computes similarity weights of all query term with respect to all tags associated with a resource as a ranking function in equation (1).

$$\theta_{Q_o,r_i} = \frac{\sum_{q_t = t_k}(wt_{Q_o,q_t} \,.\, wt_{t_k,r_i})}{\sqrt{\sum_{q_t} wt^2_{Q_o,q_t}}\sqrt{\sum_k wt^2_{t_k,r_i}}} \ldots . (1)$$

Query vector is computed by using Semantic Distance as in equation (2)

$$Wt_{Q_o,q_t} = \frac{SR_{Q_o,q_t}}{\sum_{t=0}^{n} SR_{Q_o,q_t}} = \frac{\text{Semantic Weight of qt} \in Q}{\text{Semantic Weights of all terms in Q says n terms}} \ldots (2)$$

And the vector length (magnitude) of expanded query Q is defined is given as by equation (3):

$$|Q| = \sqrt{\sum_{qt} Wt^2_{Q_o,q_t}} \ldots .. (3)$$

Resource vector is computed as semantic relevance of query term Q_o and each tag t_k associated with resource r_i with equation (4) by utilizing equation (2).

$$SR_{Q_o,t_k} = \frac{Count_{t_k}}{|d|_{Q_o,t_k} + 1} = \frac{\text{Tag Frequency}}{\text{Sematnic Weights of tags and query terms}}$$

The normalized form is:

$$Wt_{t_k r_i} = \frac{SR_{Q_o,t_k}}{\sum_{k=1}^{m} SR_{Q_o,t_k}} = \frac{\text{Semantic Weight of } t_k \text{ associated with } r_i}{\text{Semantic Weights of all tags (say m) associated with } r_i} \ldots . (4)$$

Equation (5) defines the resource vector$r_i \in$ R for equation (1) as:

$$|r_i| = \sqrt{\sum_k w\, t^2_{t_k,r_i}} \ldots .. (5)$$

In similar way A. Thukral et al [58] proposed Social Semantic Relevance (S2R) approach based on Vector Space Model which exploits relationship between tags and pre-existing semantics knowledge associated with tagged resources.

3.4.3 Agent Assisted Based Approach

G. Fenza et al [59] proposed the idea of agent-assisted tagging which aims is to assist users not only in tagging activity but in suggesting relevant resources as well. A plug-in-application has been programmed using Delicious APIs based on JADE platform for browser enriched with three types of agents:

a) The user agent Parses resources a user currently explores. After analyzing the $*$ idf , with confidence threshold 0.6, candidate words are presented to the user for tagging activity.
b) The FFCA agent (Fuzzy Formal Concept Analysis) maintains a cross-table matrix of (resources × tags) the intersection of which contains the relevance of the tag to associated resource having a value between 0 and 1.
c) The lattice agent makes a formal concept lattice tree from fuzzy formal context matrix for corresponding resources with edges having the weighting score for recommendation and ranking web resources.

3.5 Cluster Based Approaches

The clustering (unsupervised learning) is the process of organizing objects into groups having similarity with respect to some common properties while classification is considered as the supervised learning which the task of identifying to which cluster a new object belongs to, based on training set. Clustering has been adopting to exploit different aspects of the social web in order to re-order search result so that to facilitate users with the top relevant documents on the top position in decreasing order of their relevancy. Some of the technique that is based on cluster, in this regard, is briefly discussed below.

3.5.1 Tag Clustering Through Association Rules Approach

Y. Zhou et al [60] proposed tag-clustering approach based on Association Rule Mining operates in using the following steps.

a) Tag Clustering: Tags are clustered into different concepts using tags graph based on association rule mining where edge between two tags is signified with a similarity weight $W_{t_i t_j} = \text{Conf}(t_i \rightarrow t_j)$. The similarity between two clusters A and B is calculated by the following equation (1).

$$\text{sim}(A, B) = \frac{\text{cut}(A,B)}{|A|} + \frac{\text{cut}(A,B)}{|B|} \ldots . (1)$$

Whereas $|A|$ represents the number of nodes in cluster A and B, $\text{cut}(A, B)$ is the sum of cross/cut edges weights of overlapping tags between cluster A and B by equation (2)

$$\text{cut}(A, B) = \sum_{t_i \in A, t_j \in B} W_{t_i t_j \in (\text{cut edges})} \ldots . (2)$$

b) Similarity Function between resource r and concept C is given by equation (3):

$$\text{sim}(r, C) = \frac{(\sum_{t \in (r \cap C)} W(t,C))^2}{\sum_{t \in C} W(t,C) * \sum_{t \in r} w(t,r)} = \frac{\text{Sum of weights of overlapping tags in concept C}}{\text{sum of weights in C} \times \text{sum of weights of tags in r}} \ldots . (3)$$

Defines similarity between resource r and concepts C with respect to ranked tags weightages, where $W(t, r)$ represent weights of tags associated with a resource r as$W(t, r) = W(t, C) \backslash t \in C$ or $0 \backslash t \notin C$. Equation (4) calculates the concept tag's weights associated with a concept C. The cohesion is defined for a tag is the number of links (edges) with other tags in the same concept. While Inv. Coup is the number of links a tag t has with other tags not in the same concept (other external tags).

$$\text{weight}(t, C) = \begin{cases} 0 & \text{if } t \notin C \\ \text{cohesion}\left(\sum_{v \in C} W_{t,v}\right) * \text{Inv. Coup}(2^{-\sum_{u \in C} W_{t,u}}) & \text{if } t \in C \end{cases} \ldots . (4)$$

3.5.2 Web Pages and Tag Clustering (WTC)

C. Zhao [61] proposed Web pages and Tag Clustering algorithm (WTC) computed in two step, first web pages and tags are clustered using hyper graph spectral clustering algorithm. Secondly coverage rate between resources and tags rate for ranking web resources.

WorldNet is used to transform all worlds, of retrieved documents, into noun form to create a set of worlds which expresses web page more precisely. Tags and web pages are clustered using hyper-graph spectral clustering results in (a) tag clusters (TCs) and (b) web pages clusters (PCs). Each TC represent a set of web pages (PS) which contains all the pages which at least contains one tag form the corresponding TC as content word, at same pattern every page cluster (PC) represent tag set (TS) which includes all the tags which were at least associated with one page in the corresponding (PC). The similarity between documents is computed by the following equation (1).

$$\mathrm{Sim}(d_i, d_j) = \frac{1}{2}\frac{|d_i \cap d_j|}{|d_i \cup d_j|} + \frac{1}{2}\frac{|d_i \cup d_j|}{|d_i| + |d_j|} \ldots . (1)$$

After this the largest ten couples of TC and TS are considered for the retrieval of web page sets PSs and web page clusters PCs. The coverage rate between a web page cluster and web page set defines the ranking measures using equation (2).

$$\mathrm{Cov}(C_i \cap D_j) = \frac{|C_i \cap D_j|}{|C_i \cup D_j|} \ldots . (2)$$

Where C_i and D_j represent web page cluster and web page sets respectively. Query is matched against the TSs and TCs and pick up those which contains the query tag. The similarity among the TSs and TCs are computed and top ten largest TSs and TCs are considered for which the corresponding PCs and PSs are selected. The coverage between the PS and PC is computed and the couple of PS and PC is return to user having high coverage rate.

3.5.3 Similarity based Tag Clustering with Tag Frequency

S. Niwa et al [62] proposed tag clustering based on the cosine similarity among tags. Parent tags of all tags are chosen which is highly related to that tag and in this way each cluster represent a particular topic.The Personalization factor combines the recommendation points of each page within a cluster with the affinity score between users and tag clusters by equation (1).

$$\mathrm{point}(U, D) = \sum_{C_i \in \text{ CLASTERS}} \mathrm{rel}(U, C_i) \times \mathrm{point}\,(C_i, D) \ldots (1)$$

Where first part of equation (1) defines relation between users U and tags T as $\mathrm{rel}(U, C,) = \sum_{T_i \in C} \mathrm{rel}(U, T_i)$where

$$\mathrm{rel}(U, T) = \sum_{D_i \in \text{ bookmarks (U)}} \mathrm{rel}(D_i, T) \text{ with } \mathrm{rel}(D_i, T) = \mathrm{NTF}(D, T) \times \mathrm{IDF}(T)$$

The second part is the recommendation value of each page for a cluster. The recommendation point between each web page and each cluster is calculated by summing the weighted frequency score each tags within a cluster using equation (2).

$$\mathrm{point}(C, D) = \sum_{T_i \in C} w(D, T_i) \qquad (2)$$

3.5.4 Hierarchal Agglomerative Clustering Approach

A. Shepitsen et al [63] proposed for the hybrid approach of Personalized Information recommendation through Vector Space Model (VSM) with Hierarchical Agglomerative Clustering (HAC) algorithm by equation (1).

$$\theta(u, q, r) = \mathrm{Sim}(q, r) * I(u, r) = (\mathrm{VSM} * \text{Personaliztion Factor}) \qquad (1)$$

The first factor $\mathrm{Sim}(r, q)$ calculates for each document r with respect to query q by equation (2).

$$\cos(q, r) = \frac{\mathrm{tf}(q,r)}{\sqrt{\sum_{t \in T} \mathrm{tf}(t,r)^2}} \qquad (2)$$

Furthermore Agglomerative Hierarchal Clustering is used to clusters tags being examined. The distance between two clusters is the distance between their centroids. The division coefficient is taken high so that to make small clusters with similar tags. To calculate user's interest for personalization as defined by the second parameter is given in equation (3).

$$I(u, r) = \sum_{c \in C} \mathrm{uc\_w}(u, c) * \mathrm{rc\_w}(r, c) \qquad (3) \text{ where}$$

$$\mathrm{uc}_{w(u,c)} = \frac{\text{Numer of times r is annotated with a tag from a cluster c}}{\text{total number of annotations by user u}}$$

$$\mathrm{rc}_{w(r,c)} = \frac{\text{Number of times the resource r is annotated with a tag form cluster c}}{\text{total number of times the resource r is annotated}}$$

3.5.5 Semantic Tag Clustering Search

D. Vandic et al [64] proposed the idea of Semantic Tag Clustering Search (STCS) for sorting web documents which is based [65], which is further based on [66]. The sorting formula based on cosine similarity by using equation (1).

$$g(q, r) = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{m}\sum_{i=i}^{m} \cos(q_i, r_j)\right) \quad (1)$$

The result of resources been returned be solved by calculating the similarity between query tag and resources. Where $q_i \in q = \{q_1, q_2, \ldots, q_m\}$ and tags associated with resource $r_j \in r = \{r_1, r_2, \ldots, r_n\}$. Non-hierarchal clustering technique is used with angular similarity based on tag co-occurrence for semantic relatedness and levenshtein similarity to detect string similarity to avoid syntactic variations using equation (2).

$$\omega_{ij} = z_{ij} \times (1 - \widetilde{\mathrm{lev}}_{ij}) + (1 - z_{ij}) \times \cos(\mathrm{vector}(i), \mathrm{vector}(j)) \quad \text{where}$$

$$z_{ij} = \frac{\max(\text{length}(t_i).\ \text{length}(t_j))}{\text{length}(t_k)} \in [0,1] \quad (2)$$

3.6 Language Model Based Approaches

M. Ponte et al [67] proposed language model for information retrieval which is also used for social information retrieval systems. Where each document is considered as language model viewed as topic model; the probability of generating the query term by each language model is calculated and ranked with respect to those probabilities shown in equation (1).

$P(q\backslash D) = \prod_{i=1}^{k} P(q_i\backslash D)$ Using Bayes' rules

$$P(d\backslash D, u) = \frac{P(q\backslash D,u)P(D\backslash u)}{P(q\backslash u)} \quad (1)$$

S. Xu et al [68] proposed Language Annotation Model (LAM) which exploits the contents (Contents Model) and annotations (Annotation Model) set of a resource. The content model id further divided into topic cluster model and content unigram model by using the relationship among annotations and documents. The Annotation model is divided into Annotation Unigram Model as word similarity independency model and Annotation Dependency model as word similarity dependency model. The Language Annotation Model Function is giving in equation (2):

$$P(q\backslash d) = \{\text{Content Unigram Model} + \text{Topic Cluster Model} + \text{Annotatin Unigram Model} + \text{Annotation Dependancy Model}\}$$

$$P(q\backslash d) = \prod_{i=1}^{m}\{\lambda_{cum} P_{cum}(q_i\backslash d) + \lambda_{tcm} P_{tcm}(q_i\backslash d) + \lambda_{aum} P_{aum}(q_i\backslash d) + \lambda_{adm} P_{adm}(q_i\backslash d_m)\} \quad (2)$$

X. Wu et al [69] proposed the language model by using the tag assignment as conceptual space among tags, resources and users by using equation (3).

$$P(d\backslash t) = \sum_{\mu}^{D} P(d\backslash c_\mu) P(c_\mu\backslash d) \quad (3)$$

M. Harvey et al [70] proposed ranking model based on Latent Dirichlet Allocation Model (LDA) for social tagging model. The LDA model is modified for tagging topic model (TTM) to formulate the folksonomy structure into it. The ranking functions in case of LDA, and TTM1 and TTM2 are defined by equations (4), (5), (6), (7).

$$P(d\backslash q) \propto P(d).P(q\backslash d) = \frac{N_d)}{N}\prod_{w\in q} P(\omega\backslash d) = \frac{N_d)}{N}\prod_{w\in q}\sum_z P(\omega\backslash z)\,P(z\backslash d) \quad (4)$$

$$P(d\backslash q,u) \propto P(d\backslash u).P(q\backslash d,u) = P(d\backslash u)\prod_{w\in q} P(\omega\backslash d,u) \quad (5)$$

$$P(d\backslash u) = P(d)\ \sum_z \frac{P(z\backslash d)P(z\backslash d)^{\pi_u}}{P(z)} \text{ And } P(\omega\backslash d,u) = \frac{\sum_z P(\omega\backslash z)P(z\backslash d)P(z\backslash u)^{\pi_u}P(z)^{-1}}{\sum_z P(z\backslash d)P(z\backslash u)^{\pi_u}P(z)^{-1}} \quad (6)$$

$$P(d\backslash u) = P(d)\ \sum_z \frac{P(d\backslash z)P(z\backslash u)^{\pi_u}}{P(z)} \text{ And } P(\omega\backslash d,u) = \frac{\sum_z P(\omega\backslash z)P(d\backslash z)P(z\backslash u)^{\pi_u}}{\sum_z P(d\backslash z)P(z\backslash u)^{\pi_u}} \quad (7)$$

Z. Zhou et al [71] proposed language model based on risk minimization retrieval model in the context of web tagged document in the following two ways:

a) Language model is expanded with user interests where content based topic is similar to tag-based categories.

$$\lambda_1.\Delta\{P(q\backslash D), P(w\backslash q)\} + \lambda_2.\Delta\{P(z_w\backslash D), P(z_w\backslash q)\} + (1-\lambda_1-\lambda_2).\Delta\{P(i\backslash U_D), P(i\backslash U_q)\}$$

b) Language model with user interests:

$$\lambda_1.\Delta\{P(w\backslash D), P(w\backslash q)\} + \lambda_2.\Delta\{P(z_w\backslash D), P(z_w\backslash q)\} + (1-\lambda_1-\lambda_2).\Delta\{P(i\backslash U_D), P(i\backslash U_q)\}$$

2. CONCLUSION AND FUTURE SUGGESTIONS

The research problem that has been generated by the collaborative tagging has gaining popularity among different research communities and academic institutions. Many research publications have contributed in the area but with the increasing volume and relational complexity of social web paradigm makes it more favorable for scientists to work in. This survey has brought major contributions toward search result ranking of bookmarked resources in social bookmarking systems. Different properties of folksonomy model have been exploited for ranking tagged web documents against query-tag. The adopted PageRank algorithm is followed to work with the structure of folksonomy systems in order to construct graphs for random surfer to reinforce weights from node to node for ranking purposes. FolkRank, SocialPageRank, GRank, GFolkRank, RecommendationPageRank and Social HITS techniques are well known examples. Social Semantics Web technologies are also being used in combination with graph and textual techniques to propose solutions for the research problem.

In future an extensive work should be dedicated to work on developing a SocialSeachEngine like Google search engine which will not only make use of the social ranking algorithms which will not traverse tagged resources but also perform indexing, consider social factors and relations among the entities to recommend and rank resources in Social Bookmarking Environments.

REFERENCES

[1]. A. Hotho, R. Jaschke, C. Schmitz, G. Stumme, Information Retrieval in Folksonomies: Searching and Ranking. Proc. of the 3rd European Semantic Web conference. Montrnegro. , 2007,411-426.
[2]. Http://www.delicious.com/about.
[3]. S. Brin, L. Page, (1998). The Anatomy of a Large-Scale Hyper-textual Web Search Web Search Engine. Proceeding of the 7th International Conference on World Wide Web. Brisbane, Australia. 107-117.
[4]. J. Kleinberg, (1999). Authoritative Sources in Hyperlinked Environments. Journal of the ACM. 604-32.
[5]. G. Jeh and J. Widom. SimRank: a measure of structural-context similarity. In KDD'02: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 538-543. ACM Press, 2002
[6]. R.Lempel, S.Moran, The stochastic approach for link-structure analysys (SALSA) and the TKC effect,
[7]. Proceedings of the 9th International World Wide web Conference, 2000
[8]. Thomas Vander Wal, (2007), Folksonomy Coinage and Definition, 604-611.
[9]. A. Hotho, R. Jaschke, C. Schmitz and G. Stumme, (2006), Bibsonomy: A Social Bookmark and Publication Sharing System. Proceeding of the First Conceptual Sturctures Tool Interoperability Workshop, pages 87-102.
[10]. A. Mathes (2004), Folksonomies - cooperative classification and communication through shared metadata, Graduate School of Library and Information Science
[11]. H. Halpin, V. Robu, H. Shepherd (2007), The Complex Dynamics of Collaborative Tagging, Proc. International Conference on World Wide Web, ACM Press
[12]. X. Li, X. Guo and Y. Zhao (2008), Tag-based Social Interest Discovery, Proceeding of the 17th International Conference on WWW, China, ACM press, P 675-68
[13]. A. Hotho, R. Jaschke, C. Schmitz and G. Stumme (2006), Emergent Semantics in BibSonomy, Informatics vol 94(2).
[14]. S. Golden and B. A. Hubernam (2006). The Structure of Collaborative Tagging Systems. Journal of Information Sciences (accepted).
[15]. T. Hammond, T. Hannay, B. Lund, J. Scott (2005). Social Bookmarking Tools (I): A General Review. D-Lib Magazine, Vol 11.
[16]. B. Lund, T. Hammond, M. Flack, T. Hannay, (2005). Social Bookmarking Tools (II): A Case Study- Connotea. D-Lib Magazine, Vol 11.
[17]. D. Millen, J. Feinberg, and B. Kerr (2005). Social bookmarking in the enterprise. The ACM Queue, 3(9). 28-35.
[18]. R. Wetzder, C. Zimmermann, and C. Bauckhage, (2008). Analyzing Social Bookmarking Systems: A Delicious Cookbook. In Mining Social Data (MSoDa) Workshop (ECAI 08)
[19]. N. J. Deka, D. Deka (2012), Tagging and Social Bookmarking: Tools of Library Services in the Digital Ear, In 8th Convention PLANNER-2012. P 102-108
[20]. L. L. Barners (2011), Social Bookmarking Sites: A Review. Journal of Collaborative Librarianship Vol 3(3), P 180-182
[21]. Fabio Cevasco (2006), Review of ten popular social bookmarking services
[22]. C. P. Lomas (2005), 7 Things You Should Know About Social Bookmarking, Education Learning Initiative, Advance Learning through IT Innovation (NLII).
[23]. T. M. Farwell, R. D. Waters (2010), Exploring the use of Social Bookmarking Technology in Education: an Analysis of Student Experience using a course Specific Delicious.com Account, Journal of Online Learning and Teaching, Vol: 6(2), P 398-408
[24]. E. Novak, R. Razzouk, T. E. Johnson (2012). The Education Role of Social Annotation Tools in Higher Education: A Literature Review, Vol 15(1). P 39-49
[25]. C. S. Redden (2010), Social Bookmarking in Academic Libraries: Trends and Applications, The Journal of Academic Liberian ship, Vol 36(3), P 219-227
[26]. M. L. Rethleftsen (2007). Tags help make libraries Del.icio.us. Library Journal, 132 (Sept. 2007), P 26-28
[27]. P. Heymann, G. Koutrika, and H. Garcia-Mlina (2008). Can social Bookmarking Improve Web Search? In WSDM '08: Preceding of the international Conference on Web search and web data mining, pages 195-206
[28]. B. Krause, A. Hotho, G. Stumme, (2008). A Comparison of Social Bookmarking with Traditional Search. ECIR'2008. 101-113
[29]. M. G. Noll, C. Meinel (2007). Web Search Personalization via Social Bookmarking and Tagging. In proceeding of ISWC 2007.
[30]. S. Xui, S. Bao, B. Fei, Z. Su, Y. Yu (2008). Exploring Folksonomy for Personalized Search. SIGIR 08 Singapore. P 155-162
[31]. D. Vallet, I. Cantador, J. M. Mose (2008), Personalized Web Search With Folksonomy based User and Document Profiles.
[32]. I. Cantador et al (2010). Content based Recommendation in Social Tagging Systems. RecSys 2010, Barcelona Spain.
[33]. Y. Cai, Q. Li (2010), Personalized Search by Tag-based User Profile and Resources Profile in Collaborative Tagging Systems. CIKM 2010, Canada, P 969-978
[34]. P. Wu, Zi-Ke Zhang (2010). Enhancing Personalized Recommendations on weighted Social Tagging Networks, Physics Procedia 3(5): P 1877-1885.
[35]. H. N. Kim, M. Rawashdeh, A. Alghamdi, A. El Saddik (2012), Folksonomy-based Personalized Search and Ranking in Social Media Services. Journal of Information System vol 37, P 61-76
[36]. H. S. Khalifa (2008). CoolRank: A Social Solution for Ranking Bookmarked Web Resources. IEEE 2008, P 208-210
[37]. V. Zanardi and L. Capra (2008). Social Ranking: Uncovering Relevant content Using Tag-based Recommender Systems. In RecSys, Switzerland
[38]. Worasit Choochaiwattana, Michael B. Spring (2009). Applying Social Annotations to Retrieve and Re-rank Web Resources. International Conference on Information Management and Engineering. pp.215-219
[39]. Daly E. M (2009), Harnessing Wisdom of The Crowds Dynamics for Time-Dependent Reputation and Ranking, International Conference on Social Network Analysis and Mining, P: 262-275
[40]. F. Durao, P. Dolog (2009). A Personalized Tag-based Recommendation in Social Web Systems. Workshop on Adaptation and Personalization for Web 2.0. P 40-49
[41]. R. Wetztker, C. Zimmermann, C. Bauckhage, S. Albayrak, (2010). I Tag, You Tag, Translating Tags for Advanced User Models
[42]. J. B. Schafer, D. Frankowski, J. Herlocker and S. Sen (2007). Collaborative Filtering Recommender Systems. The Adoptive Web. P 291-324
[43]. D. Parra-Santander, P. Brusilovsky (2009). Collaborative Filtering for Social Tagging Systems: An Experiment with Citeulike. RecSys, New York (USA).
[44]. J. Germmell, T. Schimoler, B. Mobasher, R. Burke (2012), Resource Recommendation in Social Annotation: A Linear-weighted Hybrid Approach. Journal of Computer and System Sciences. P 1160-1174.
[45]. S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su, (2007). Optimizing Web Search Using Social Annotations. In Proceeding of 16th International World Wide Web Conference (ACM Press '07). 501-510.

[46]. F. Abel, N. Henze, D. Krause and M. Kriesell (2008). A Novel Approach to Social Tagging: GroupMe. 14th International Conference on Web Information Systems and Technologies.
[47]. F. Abel, N. Henze, D. Krause and M. Kriesell (2008). On the Effect of Group Structures on Ranking Strategies in Folksonomies. In Workshop on Social Web Search and Mining at 17th Int. World Wide Web Conference.
[48]. F. Abel, M. Baldoni, C. Barogolio, N. Henze, R. Kawase, V. Patti (2010). Research Article: Leveraging Search and Content Exploration by Exploiting Context in Folksonomy Systems. Hypermedia and Multimedia, P 1-31
[49]. F. Abel, N. Henze and D. Krause (2008). Analyzing Ranking Algorithms in Folksonomy Systems. L3S Research Center, Germany.
[50]. F. Abel, M. Baldoni, C. Baroglio, N. Henze, D. Krause and V. Patti (2009). Context Based Ranking in Folksonomies. Hypertext ACM, P 209-218
[51]. P. Wen-Chih, L. Chia-Hao (2008), Ranking Web Pages From User Perspectives of Social Bookmarking Sites, International Conference on Web Intelligence and Intelligent Agent Technology, vol 1 P: 155-161
[52]. Takahashi T., Kitagawa, H. (2008), S-BITS: Social-Bookmarking Induced Topic Search, The 9th International Conference on Web-Age Information Management, P: 52-30
[53]. Tim Berners-Lee Tim, James Hendler and Ora Lassila (2001). "The Semantic Web" . Scientific American Magazine
[54]. F. Manola, E. Miller (2004). RDF primer, W3C Recommendation
[55]. D. L. McGuiness, F. Harmelen (2004), OWL Web Ontology Language Overview. W3C Recommendation.
[56]. F. Abel, M. Frank, N. Henza, D. Krause, D. Plappert and P. Siehndel (2010), GroupMe: Where Semantic Web meets Web 2.0.
[57]. B. pedi, H. Banati, A. Thukral, (2010), Social Semantic Retrieval and Ranking of eResources, Preceding of the 2010 International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom), p: 343-345
[58]. P. Bedi, A. Thukral, H. Banati (2012), Focused Crawling of tagged Web Resources Using Ontology, Journal of Computers and Electrical Engineering,
[59]. A. Thukral, H. Banati, P. Bedi (2011). Ranking Tagged Resources Using Social Semantic Relevance. International Journal of Information Retrieval, Vol 1(3). P 15-34
[60]. G. Fenza, V. Loia, S. Senatore (2011), Agent-assisted Tagging aimed at Folksonomy-Based Information Retrieval, Intelligent Agent 2011 IEEE , Italy, p: 1-8
[61]. Y. Zhou (2009), Searching and Clustering on Social Tagging Sites, International Conference on Semantics, Knowledge and Grids. P: 99-105.
[62]. C. Zhao, Z. Zhang, H. Li, X. Xie, (2011), A Search Result Ranking Algorithm Based on Web Pages and Tags Clustering, International Conference on Computer Science and Automation Engineering (CSAE), Vol 4, P: 609-614
[63]. S. Niwa, T. Doi, S. Honiden (2006). Web page recommender system based on folksonomy mining. Information Processing Society of Japan (IPSJ) Journal, 47(5): 1382-1392.
[64]. A. Shepitsen, J. Germmell, B. Mobasher (2008). Personalized Recommendation in Social Tagging Systems using Hierarchical Clustering, In Proceedings of the 2008 ACM Conference on Recommender Systems (RecSys 2008) (October 2008), pp. 259-266,
[65]. D. Vandic, J. Van Dam, F. Hogenboom (2011). A Semantic Clustering Based Approach for Searching and Browsing Tag Spaces. ACM SAC'11. Taiwan
[66]. J. W. Van Dam, D. Vandir, F. Hogenboom, F. Fransincar (2010). Searching and Browsing Tag Spaces Using the Semantic Tag Clustering Search Framework. 4th IEEE International Conference on Semantic Computing (ICSC 2010), p 436-439
[67]. L. Specia, E. Motta (2007). Integrating Folksonomies with the Semantic Web. 4th European Semantic Web Conference (ESWC 2007). LNCS P 503-517
[68]. M. Ponte and W.B. Croft (1998). A Language Model Approach to Information Retrieval. Proceeding of the 1st ACM SIGIR 1998. P: 275-281.
[69]. S. Xu, S. Bao, Y. Yo (2007). Using Social Annotations to Improve Language Model for Information Retrieval. In CIKM'07 ACM. P 1003-1006.
[70]. X. Wu, L. Zhong and Y. Yo (2006). Exploring Social Annotations for the Semantic Web. In WWW ACM Scotland.
[71]. M. Harvey, I. Rthven, M. J. Carman (2011). Improving Social Bookmarking Search Using Personalized Latent Variable Language Models. WSDM'11 ACM, China. P 485-494
[72]. Z. Zhou (2011), Social Information Retrieval Based on User Interesting Mining, Journal of Computational Information Systems 7:4 (2011) 1373-1379