Xie HR, Li Q, Cai Y. Community-aware resource profiling for personalized search in folksonomy. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 27(3): 599–610 May 2012. DOI 10.1007/s11390-012-1247-7

Community-Aware Resource Profiling for Personalized Search in Folksonomy

Hao-Ran Xie[1] (谢浩然), Qing Li[1] (李　青), *Senior Member*, *CCF*, *IEEE*, and Yi Cai[2,*] (蔡　毅), *Member*, *CCF*

[1] *Department of Computer Science, City University of Hong Kong, Hong Kong, China*

[2] *School of Software Engineering, South China University of Technology, Guangzhou 510006, China*

E-mail: hrxie2@student.cityu.edu.hk; itqli@cityu.edu.hk; ycai@scut.edu.cn

Received September 1, 2011; revised January 20, 2012.

Abstract In recent years, there is a fast proliferation of collaborative tagging (a.k.a. folksonomy) systems in Web 2.0 communities. With the increasingly large amount of data, how to assist users in searching their interested resources by utilizing these semantic tags becomes a crucial problem. Collaborative tagging systems provide an environment for users to annotate resources, and most users give annotations according to their perspectives or feelings. However, users may have different perspectives or feelings on resources, e.g., some of them may share similar perspectives yet have a conflict with others. Thus, modeling the profile of a resource based on tags given by all users who have annotated the resource is neither suitable nor reasonable. We propose, to tackle this problem in this paper, a community-aware approach to constructing resource profiles via social filtering. In order to discover user communities, three different strategies are devised and discussed. Moreover, we present a personalized search approach by combining a switching fusion method and a revised needs-relevance function, to optimize personalized resources ranking based on user preferences and user issued query. We conduct experiments on a collected real life dataset by comparing the performance of our proposed approach and baseline methods. The experimental results verify our observations and effectiveness of proposed method.

Keywords tagging, personalized search, user community, social filtering

1 Introduction

In recent years, there is a fast proliferation of collaborative tagging (a.k.a. folksonomy) systems in Web 2.0 communities. Tagging becomes an important way of indexing and organizing user interested resources. The rich semantics from user-generated tags have been utilized in various applications such as bookmark collection (Del.icio.us[1]), movie recommendation (Movielens[2]) and image sharing (Flickr[3]). With the ever increasing amount of user-generated tags and resources, how to assist users to find their interested resources is one of the most important issues for these applications.

One main stream of solutions to this problem is to construct user profiles and resource profiles derived from folksonomies in order to facilitate personalized search[4-7]. For these existing works, the profile of a resource is constructed based on tags given by all users who have annotated the resource. Collaborative tagging systems provide a way for users to annotate resources, and most users give annotations according to their perspectives or feelings. However, users may have different perspectives or feelings on resources. For example, some of them may share similar perspectives (or feelings) yet have a conflict with others.

To achieve personalization in the process of search, we need to describe resources by their profiles as close as to a user's real perspective or feeling on them. In other words, tags in the profile of a resource should be close or similar to those which have been used by users in annotating the resource. If the profile of a resource is constructed by tags given by all users who have annotated the resource, however, it may distort the description of the resource from an individual user's perspective. Let us take a look at the following example.

Regular Paper

This work was supported by the Research Grants Council of Hong Kong SAR under Grant No. CityU 117608, a strategic research grant from City University of Hong Kong under Project No. 7002606, Foundation for Distinguished Young Talents in Higher Education of Guangdong Province of China under Grant No. LYM11019, the Natural Science Foundation of Guangdong Province of China under Grant No. S2011040002222, and the Fundamental Research Funds for the Central Universities of South China University of Technology under Grant No. 2012ZM0077.

This paper is an extended version of our previous conference paper[8].

*Corresponding Author

Example 1. Suppose that there are three users Alice, Bob and Carol. Bob is a fan of light food, while Alice and Carol like spicy food. As shown in Fig.1, Bob gives tags on the recipe "Kung Pao Chicken" as "hot" and "salty" since the recipe is spicy to Bob. Alice gives the tags "mild" and "light" to the recipe because she likes more spicy food in her daily life. If we adopt the tags given by Bob and Alice on the recipe to construct its profile (suppose we use term frequency to construct the profile), then the description of the recipe will have inconsistence, with the profile being the following:

$$\boldsymbol{R} = (hot : 1, salty : 1, mild : 1, light : 1).$$

When Carol (who shares the similar taste as Alice) tries to search some spicy recipes and input the keyword "hot", the recipe "Kung Pao Chicken" may also be returned even though Carol regards this recipe as quite light. This is because the recipe profile matches the keyword "hot" even though Carol does not think it is a spicy food. The problem is due to that the tag "hot" is not the same or similar to Carol's real feeling on "Kung Pao Chicken". In other words, Carol and Bob have different or conflicting feeling on this recipe. So it is not reasonable to adopt the resource profile constructed based on all users' tags to achieve personalized search for an individual user.

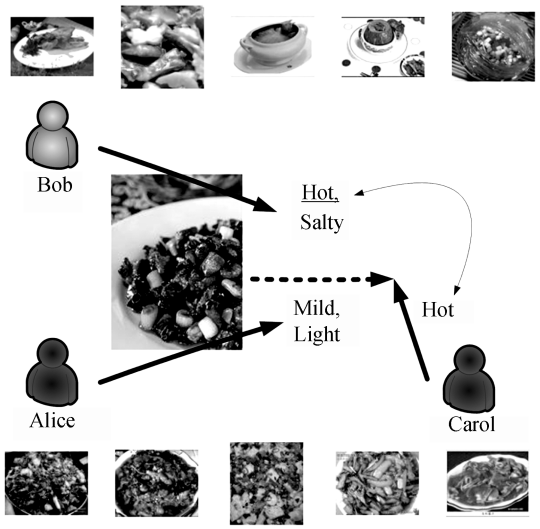

Fig.1. Example of conflicting opinions from different users in folksonomies.

Thus, we consider that modeling the profile of a resource based on tags given by all users who have annotated the resource is neither suitable nor reasonable. In this paper, we propose a community-aware approach to constructing resource profiles via social filtering. Since people who share similar interests and have the common perspectives on resources are likely to be in some forms of community, we can utilize these "close views" from the community to model resource profiles in a more precise way. The contributions of this paper are listed below.

• We propose three different strategies to establish the relationships between resources and communities; so as relationships play an essential role in discovering user community.

• We propose a social filtering method to distinguish two kinds of resource profiles (social filtering resource profile versus collective resource profile) based on the characters of users.

• By utilizing the resource profiles, we devise a personalized search approach which combines our proposed switching fusion method with a revised needs-relevance function, so as to optimize personalized resources ranking based on user preferences and user issued query.

• We conduct experiments on the real-life dataset by comparing the performance of our proposed approach and baseline methods. The experimental results verify our observations.

The remainder of this paper is structured as follows. In Section 2, we review some relevant work. We describe our model for constructing user profiles, user communities and two kinds of resource profiles in folksonomies in Section 3. A personalized search approach to optimizing resources ranking based on user preference and user issued query is described in Section 4. We conduct experiments and analyze their results in Section 5. Finally, we conclude our work and give possible future research directions in Section 6.

2 Related Work

In this section, we review some relevant work of collaborative tagging and personalized search in the folksonomies environment.

2.1 Collaborative Tagging

Recent research in collaborative tagging can be categorized into two aspects. One is on investigating and analyzing the characteristics of user generated tags. In [9], tag usage patterns and user behavior in tagging are studied by Golder and Huberman. In order to discover valuable tags for search, Bischoff *et al.*[10] did a survey on some real tagging datasets. Manish *et al.*[11] did a comprehensive survey on various features of social tagging data and techniques. The other is to discover some features such as link structure, semantic similarities in the folksonomies for various applications. Bao *et al.*[12] proposed two novel algorithms, which are SocialSimRank (SSR) and SocialPageRank (SPR), to incorporate benefits from social annotations in order to facilitate web search. In [13], three (naive, co-occurrence and adaptive) approaches to constructing the tag-based

profile and their comparison were studied. A survey on different metrics to measure the semantic similarity between tag-based profile was done by Markines *et al.*[14].

2.2 User Community

Through ways of combining community information with content-based ranking, a community-aware search engine was proposed by Almeida and Almeida[15]. Park *et al.*[16] applied community popularity to PageRank and got more general form. In [17], Smyth utilized Hit-Matrix from a community to support personalizing web search through a collaborative way. Meanwhile, how to enhance expertise retrieval by community-aware strategies has been studied by Deng *et al.*[18] In [19], the explicit user community (group) information was utilized in different kinds of search tasks.

2.3 Personalized Search

Personalized search is a crucial way to bridge the large gap between how well search engines could perform if they were to tailor results to individuals, and how well they currently perform by returning results designed to satisfy everyone[20]. By adopting collaborative filtering method, Liu and Yang[21] devised a personalized approach EigenRank to recommend items according to user preferences. In [22], an interest-based personalized search framework, which maps user interests onto a group of categories in the Open Directory Project (ODP), was proposed to categorize and personalize search results. Carmel *et al.*[23] investigated personalized social search based on the users' social relation. In addition, topic model[24], user web log[25-27], online social activities[28], concept relations[29-31] and user communities[17,19,32] have been exploited as the indicator to facilitate personalized search. A performance comparison among various personalized strategies was studies by Dou *et al.*[33]

2.4 Personalized Search in Folksonomies

There are some existing studies on utilizing resource profile and user profile to facilitate personalized search in folksonomies. Noll and Meinel[5] proposed a term frequency (TF) profiles to discover related tags for users and resources, so that personalized ranking is provided. The later studies follow the term frequency-inverse document frequency (TF-IDF), Best Matching 25 (BM25)[7] and their hybrid[6] paradigms. In [34], TF-IDF was combined with the user and resource profiles along with positions of tags, by considering two kinds of sources. Furthermore, in our earlier work[4], we proposed a normalized term frequency (NTF) to model user and resource profiles and compare it with previous methods. However, these methods use tags from all users to build a resource profile, so that the conflicting opinions from user annotations are not screened out but maybe imported into the resource profile.

3 Resource and User Modeling

3.1 User Profiling

In a collaborative tagging system, the tags which are used by a user to annotate resources can reflect this user's preference to some extend. By following this observation and existing forms of user profiles in collaborative tagging systems, we define a user profile as follows.

Definition 1. *A user profile of user i, denoted by $\boldsymbol{U}_i$, is a vector of tag:value pairs, i.e.,*

$$\boldsymbol{U}_i = (t_{i,1} : v_{i,1}, t_{i,2} : v_{i,2}, \ldots, t_{i,n} : v_{i,n}),$$

where $t_{i,x}$ is a tag annotating some resource by user i, n is the total number of tags used by user i, $v_{i,x}$ is the preference degree of user i on tag $t_{i,x}$. As discussed in our previous work[4], it can be obtained by NTF as follows①:

$$v_{i,x} = \frac{N_{i,x}}{N_i}, \tag{1}$$

where $N_{i,x}$ is the number of times user i uses tag x to annotate resources, and N_i is the number of resources tagged by user i. The higher value of $v_{i,x}$ is, the more preferred (favorable) is tag x by user i.

3.2 Collective Resource Profiling

For a resource in the folksonomies, how well tag x is used to describe resource d is dependent on the possibility or proportion of users' using tag x to annotate d among all the users who have annotated d. Therefore, a resource profile can be defined as follows.

Definition 2. *A collective resource profile of a resource c, denoted by $\boldsymbol{R}_c$, is a vector of tag:value pairs*:

$$\boldsymbol{R}_c = (t_{c,1} : w_{c,1}, t_{c,2} : w_{c,2}, \ldots, t_{c,n} : w_{c,n}),$$

where $t_{c,x}$ is a tag being used to describe resource c, n is the number of tags used to describe resource c, $w_{c,x}$ is a weight value to which resource c possesses the tag (feature) $t_{c,x}$, and $w_{c,x}$ can be intuitively obtained as follows:

$$w_{c,x} = \frac{M_{c,x}}{M_c}, \tag{2}$$

where $M_{c,x}$ is the number of users using tag x to annotate resource c, and M_c is the total number of users

①The preference degree in both user and resource profiles can be obtained in other ways such as TF, TF-IDF or BM25, and we will demonstrate them in the experiments.

who use tags to annotate resource c. A higher value of $w_{c,x}$ means that tag x is more salient or representative for resource c.

3.3 Community Modeling

The purpose of the resource profiling in personalized search is to predict how a user may describe a resource based on his or her personal perspective and feeling, so as to measure how the resource matches the query and how relevant the resource is to the user. However, users may have different perspectives or feelings on resources. Some users share similar perspectives (or feelings) yet may have conflict with other users. The collaborative resource profiles are constructed from annotations by all users. Conflictive tags may be given by different users for the same resource, and unreasonable results may return when the personalized search is performing (as discussed in Example 1).

Hence, collective resource profiles which are modeled by all user tags on resources are unsuitable and unreasonable. To overcome the shortage of collective resource profiles, we need to avoid conflictive annotations from users. In real life, people usually form various communities via similar interests in both virtual and real worlds[35]. Within a community, members usually share similar interests and have close perspectives on resources. We define the set of users who share a similar perspective or feeling with each other as a community.

Definition 3. *A community, denoted by $\boldsymbol{C}_p$, is a vector of user:value pairs*:

$$\boldsymbol{C}_p = (U_{p,1} : s_{p,1}, U_{p,2} : s_{p,2}, \ldots, U_{p,n} : s_{p,n}),$$

where $U_{p,a}$ is a user who belongs to the community, n is the number of users belonging to the community, $s_{p,a}$ is the degree of membership for user a to be in the community.

We consider that a community is associated with some resources. For example, a science fiction movie community is composed of science fiction movie fans and associated with resources such as "Matrix", "Star Wars". To obtain the associated relationships, we propose the following three strategies.

Subjective Derivation. This strategy requires users to decide whether a resource is associated to a community or not without any hints from the system. For example, it may ask several people to decide whether or not "Star Wars" belongs to science fiction movie community. In other words, the wisdoms of the crowd are tended to derive a reasonable result. Based on this assumption, judgements by many people are necessary, and the degree of a resource associated to a community can be decided by their overall opinions, i.e.,

$$k_{p,i} = \sum_{n=1}^{N} \frac{opin_n(r_i, \boldsymbol{C}_p)}{N}, \tag{3}$$

where $k_{p,i}$ is the membership degree for resource r_i to community p, N is the total number of human opinions, $opin_n(r_i, \boldsymbol{C}_p)$ is an opinion result function. "1" means that resource r_i belongs to community p in the opinion; otherwise, "0" is given. As to be shown by our experiments, this strategy is more accurate than others. However, it is very time-consuming and requires too much human effort to acquire reasonable associated relationships.

Heuristic Propagation. To reduce the human workload required by the subjective derivation strategy, a heuristic strategy is proposed to discover the relationships between the resource and communities. This strategy can be further divided into two sub-processes. First, a small subset A of relationships is measured by subjective derivation as above. Then, the remaining relationships can be obtained through propagating resources in A by considering the similarity between two collective resource profiles, as follows:

$$k_{p,i} = \sum_{\forall j \in S} \frac{sim(r_i, r_j) \times k_{p,j}}{|S|}, \tag{4}$$

$$sim(r_i, r_j) = \frac{\boldsymbol{R}_i \cdot \boldsymbol{R}_j}{\mid \boldsymbol{R}_i \mid \times \mid \boldsymbol{R}_j \mid}, \tag{5}$$

where $S = \{j | r_j \in A,\ \boldsymbol{R}_i \bigcap \boldsymbol{R}_j \neq \varnothing\}$, $sim(r_i, r_j)$ is the similarity between two resources r_i and r_j based on their collective resource profiles $\boldsymbol{R}_i$ and $\boldsymbol{R}_j$ respectively, $k_{p,j}$ is the membership degree for resource r_j to community p, $|S|$ is the total number of elements in set S. This heuristic strategy is quite efficient to propagate most associated relationships through a small proportion of the original set, and we will further discuss it later.

Cluster-Based Generation. There are many existing cluster-based approaches[36-39] to discovering resources associated for a community. Since our focus is on how to utilize the community to assist personalized search, here we adopt one possible solution which is the topic model[36], so as to obtain these associated relationships among resources and communities. The graphical representation of the topic model is shown in Fig.2. In particular, the topic model we use is the Latent Dirichlet Allocation (LDA)[40], and we assume that each community corresponds to a topic. Accordingly, the generative process can be described formally as follows.

• For each topic z, draw β_z from $Dirichlet(\mu)$;

• For each tag t_{ri} in resource r, draw a topic c_{ri} from $Multimomial(\theta_{ri})$, where θ is generated from $Dirichlet(\alpha)$.

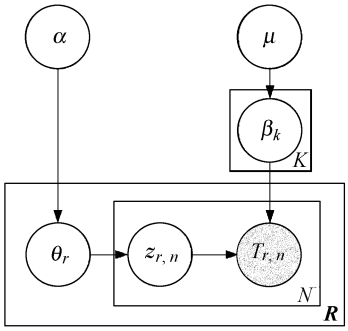

Fig.2. Graphical representation of the topic model.

The generating probability of tag t_{ri} from resource r is given below:

$$P(t_{ri}|r,\theta,\beta)=\sum_{z=1}^{K}P(t_{ri}|z,\beta_z)P(z|r,\theta_r), \qquad (6)$$

where z is a topic, t_{ri} is the i-th tag in resource r, α and μ are Dirichlet priors to multinomial distribution θ_r and β_k, respectively, N is the total number of tags in resource r, R is the number of resources, and K is the total number of topics. The parameters are estimated by Gibbs sampling to infer model parameters θ and β directly. The degree value $k_{p,j}$ of associated resource r to community p can be regarded as the topic assignment to the resource.

The relationships among users, communities and resources are illustrated in Fig.3. A user can belong to multiple communities with different degree of membership (e.g., Alice may prefer science fiction movies mostly and romantic movies secondly). Similarly, a resource can be associated with multiple communities by different degree of relevance values, e.g., movie "Matrix" is related to both science fiction and action movies. Based on user behaviors on the associated

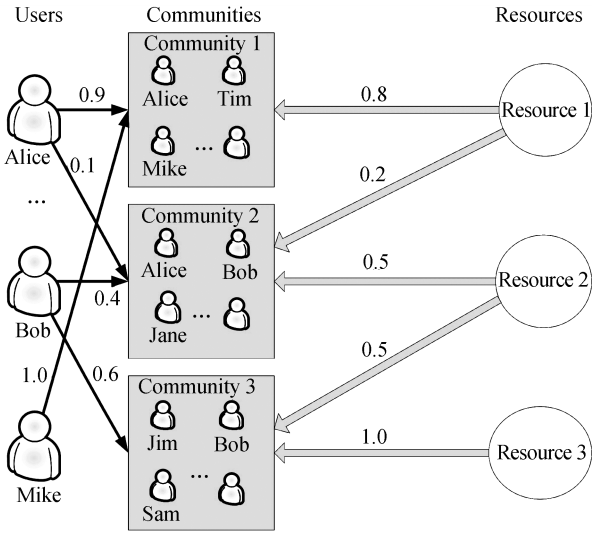

Fig.3. Relationships among users, communities and resources.

resources for a community, we make the following assumption.

Assumption 1. *For two communities p and q and user i, if* 1) *user i more frequently annotates resources from community p than from q, and* 2) *the annotated resources by user i are more relevant to community p than q, then we assume that user i has a greater membership degree to community p than to q.*

According to the above assumption, how much user i belongs to a community p depends on two factors: the possibility or proportion of his/her annotations on resources, and resource membership for community p among all his/her annotated resources. Based on Assumption 1, we can obtain the user membership degree for a specific community as follows:

$$s_{p,i}=\frac{L_{p,i}}{L_i}\times\frac{\sum_{r_j\in Ri}^{\forall j}k_{p,j}}{L_{p,i}}=\frac{\sum_{r_j\in Ri}^{\forall j}k_{p,j}}{L_i}, \qquad (7)$$

where $L_{p,i}$ is the number of annotated resources by user i in the associated resources set for community p, L_i is the number of resources annotated by user i, R_i is the set of resources user i has tagged, and $k_{p,j}$ is the membership degree for resource r_j to community p. The higher value of $s_{p,i}$, the greater degree of membership for user i to community p.

3.4 Social Filtering Resource Profiling

Within a community, users might have different degrees of memberships. Users who have the higher degrees of memberships can reflect better characteristics of the community.

In order to find the most relevant users for a community, we set a threshold η_p for selecting a high membership degree for a community. If the degree of membership for a candidate user i to community p is greater than or equal to the threshold η_p, user i is selected into the community core.

Definition 4. *The community core of a community p, denoted by $Core_p$, is a set of users in the community p whose degree of membership is greater than threshold η_p*:

$$Core_p=\{U_{p,i}|s_{p,i}>\eta_p\},$$

where $U_{p,i}$ is a user who belongs to community p, $s_{p,i}$ is the value of degree of membership to community p, η_p is the threshold for selecting most relevant users to community p.

To avoid arbitrarily determining η_p, we set η_p by using the Chebyshev Law[41] as follows:

$$\eta_p=\gamma_p-k\times\sigma_p, \qquad (8)$$

where γ_p is the average relevance degree value for all users to community p, σ_p is the standard deviation, k

is an integer and a parameter to set the size of the community core. According to the Chebyshev Law, for any integer $k > 1$, at least $(1 - \frac{1}{k^2})$ sample values are in the interval $(\gamma_p \pm k \times \sigma_p)$. That is, if we set $k = 2$, at least $\frac{3}{4}$ users are in the range of $(\gamma_p \pm 2 \times \sigma_p)$. And η_p here is to select users whose relevance degree values are in the interval $[\gamma_p - k \times \sigma_p, 1]$, thus about $\frac{3}{4} + \frac{1}{2} \times (1 - \frac{3}{4}) = \frac{7}{8}$ users will be selected into $Core_p$.

As discussed previously, it is not reasonable to use the resource profiles constructed according to all users' taggings to achieve personalized search for any particular user. However, users in a community do share similar perspectives on resources and we name such perspectives as community view. Thus, we regard that a resource profile with respect to a particular user i is based on that resource's tags given by users who come from the community cores of relevant communities for user i, and such a resource profile is named as social filtering resource profile.

Definition 5. *A social filtering resource profile of a resource c for a user i, denoted by $\boldsymbol{R}_c^i$, is a vector of tag:value pairs*:

$$\boldsymbol{R}_c^i = (t_{c,1}^i : w_{c,1}^i, t_{c,2}^i : w_{c,2}^i, \ldots, t_{c,n}^i : w_{c,n}^i),$$

where $t_{c,x}^i$ is a tag being used to describe resource c by users from community cores of user i's relevant communities, n is the number of tags used to describe resource c by users from community cores of user i's relevant communities, $w_{c,x}^i$ is to measure how much resource c possesses the tag (feature) $t_{c,x}^i$, and it can be obtained as follows:

$$w_{c,x}^i = \frac{M_{c,x}^i}{M_c^i}, \tag{9}$$

where $M_{c,x}^i$ is the number of users from community cores of user i's relevant communities using tag x to annotate resource c, and M_c^i is the total number of users from community cores of user i's relevant communities in which tags are used to annotate resource c.

4 Personalized Search

In a personalized search system, user queries and user profiles to some extent represent their information needs. One of the information needs is usually represented by the user issued query terms. For example, Bob may issue a query which contains terms "fish" and "spicy" to search recipes in which the main ingredient is fish and the taste is spicy. Because this kind of needs is specified by users explicitly, we name this kind of information needs as *explicit information needs*. In contrast, user profiles are different from explicitly specified user query terms and can be regarded as the *implicit information needs* of the users.

These two kinds of information needs should be taken into consideration when the personalized search is performing. Hence, we measure how likely a relevant resource is able to satisfy these two kinds of information needs in our personalized search approach. This approach can be divided into two sub-processes. One is information needs fusion which is to unify the explicit and implicit information needs to a unified form to represent a user's current information needs. The other is resources ranking, which is to rank the resources by the score of relevance between user fusion information needs and the corresponding social filtering resource profile.

4.1 Information Needs Fusion

We assume a user query (explicit information needs) can be represented by a vector of terms, as defined below.

Definition 6. *A query issued by user i, denoted by $\boldsymbol{Q}_i$, is a vector of terms, i.e.,*

$$\boldsymbol{Q}_i = (t_{i,1}^q : v_{i,1}^q, t_{i,2}^q : v_{i,2}^q, \ldots, t_{i,m}^q : v_{i,m}^q),$$

where $t_{i,x}^q$ is a term, m is the total number of terms in the query, $v_{i,x}^q$ is the value indicating the importance degree of $t_{i,x}^q$ for $\boldsymbol{Q}_i$. We further assume that the relationship among all terms in a query are conjunctive, e.g., a query which contains terms "fish" and "spicy" means that query issuer wants to find out resources which possess all the query terms. We consider that all terms in a query have the same importance degree. Thus, the values of all $v_{i,x}^q$ are given as 1.

The implicit information needs for a user are embodied by his/her profile, as defined in Definition 1. We aggregate query terms and user profile into fused information needs as the first prior of our approach.

Definition 7. *The fused information needs for a user i, denoted by $\boldsymbol{F}_i$, is a vector of tag:value pairs*:

$$\boldsymbol{F}_i = (t_{i,1}^f : v_{i,1}^f, t_{i,2}^f : v_{i,2}^f, \ldots, t_{i,n}^f : v_{i,n}^f), \quad \forall x, t_{x,n}^f \in \boldsymbol{Q}_i \cup \boldsymbol{U}_i,$$

where $t_{i,x}^f$ is a tag by user i, $v_{i,x}^f$ is the corresponding importance degree for $t_{i,x}^f$ to the fused information needs, and n is the total number of tags.

Intuitively, the importance degree of $t_{i,x}^f$ for user i's fused information needs can be obtained by the function below:

$$v_{i,x}^f = \delta \times v_{i,x}^q + (1 - \delta) \times v_{i,x}, \tag{10}$$

where δ is an adjusting parameter in the range of $[0, 1]$. As discussed above, we assume that a users' interests are reflected by his/her query terms. Thus, the terms in the query will get higher weighted values in the fused information needs. However, this linear combination

method has some shortage, as shown by the following example.

Example 2. User Tom has issued the following query, with his user profile given below.

$$\boldsymbol{Q}_{Tom} = (braise : 1.0,\ chicken : 1.0),$$
$$\boldsymbol{U}_{Tom} = (spicy : 0.3,\ icecream : 1.0).$$

If we use the linear fuse approach, and set δ to be 0.6, the following fused information needs vector will be obtained:

$$\boldsymbol{F}_{Tom} = (braise : 0.6,\ chicken : 0.6,\ spicy : 0.12,\ icecream : 0.4).$$

We see that icecream in the user profile $\boldsymbol{U}_{Tom}$ is included into the fused information needs vector. Hence, those recipes which only contain term "icecream" will be included in the query result. However, if we examine user intention more carefully, which is to search some "braised chicken" recipes, it is unreasonable to include "icecream" into the fused information needs vector in such a case.

According to above analysis, we can see that not all the terms in a user profile are useful to process a query issued by the user. To tackle this problem, we make the following assumption.

Assumption 2. *For each user query, not all tags/terms in the user profile are useful, but only those tags/terms appeared in the query terms in the resources are valuable to the user.*

Based on this assumption, we propose a switching fused method as follows:

$$v_{i,x}^{f} = \begin{cases} 1, & \text{if } t_{i,x} \in \boldsymbol{Q}_i, \\ v_{i,x}, & \text{if } t_{i,x} \notin \boldsymbol{Q}_i \text{ and } \exists y, c, t_{i,x} \in \boldsymbol{R}_c, \\ & t_{i,y}^{q} \in \boldsymbol{R}_c,\ t_{i,y}^{q} \in \boldsymbol{Q}_i, \\ 0, & \text{otherwise}, \end{cases} \tag{11}$$

where this piecewise function specifies three cases:

1) If the tag/term $t_{i,x}$ occurs in both the user profile and query, it will be given "1" due to its importance;

2) If the tag/term $t_{i,x}$ occurs in the user profile only and not in the query, yet it also occurs in some resource profile relevant to some query term, it will be kept and included in the fused information needs vector since this co-occurrence indicates that the tag/term should be somewhat relevant to the query;

3) Otherwise, this tag will be set as "0", and then excluded from the fused information needs vector by the operation of $\boldsymbol{F}_i - \{t_{i,x}\}$.

Let us revisit Example 2. Since "icecream" rarely appears with query terms "braised" and "chicken", by using the switching fused method, it will be excluded from the fused information needs vector. But the tag "spicy" will be kept due to that it has appeared together with the query terms "braised" or "chicken' in some recipes. So, after performing the excluding operation $\boldsymbol{F}_i - \{t_{i,x}\}$, the fused information needs vector is shown as follows:

$$\boldsymbol{F}_{Tom} = (braise : 1.0, chicken : 1.0, spicy : 0.3).$$

Consequently, we regard it as more reasonable to adopt the switching method than the linear method during the information need fusion phase.

4.2 Resource Ranking

Next, we measure how relevant a candidate resource matches to user information needs and rank the resource based on the relevant score. The relevance score of a resource to user information needs can be measured by a needs relevance function ϕ:

$$\phi : \boldsymbol{F} \times \boldsymbol{R} \to [0, 1],$$

where $\boldsymbol{F}$ is the set of fused information needs and $\boldsymbol{R}$ is the set of resources. The result of the ϕ function is the relevant score of a resource to the fused information needs. The higher the relevant score, the more relevant is the resource to the fused information needs.

Since user profiles and fused information needs are in the form of vectors, it is straightforward to measure their relevance by the cosine as follows:

$$\phi(\boldsymbol{F}_i, \boldsymbol{R}_c^i) = \frac{\boldsymbol{F}_i \cdot \boldsymbol{R}_c^i}{\mid \boldsymbol{F}_i \mid \times \mid \boldsymbol{R}_c^i \mid}. \tag{12}$$

However, the cosine measurement has a shortage itself, as the following example shows.

Example 3. User Bob has the following fused information needs:

$$\boldsymbol{F}_{Bob} = (chicken : 1.0, spicy : 0.9, pork : 0.2).$$

There are two resources c and d, and their corresponding social filtering resource profiles are as follows:

$$\boldsymbol{R}_c^{Bob} = (chicken : 0.5, spicy : 0.45, pork : 0.1),$$
$$\boldsymbol{R}_d^{Bob} = (chicken : 1.0, spicy : 1.0, pork : 0.5).$$

If we use cosine to measure the relevant score, we will obtain the following results:

$$\phi(\boldsymbol{F}_{Bob}, \boldsymbol{R}_c^{Bob}) > \phi(\boldsymbol{F}_{Bob}, \boldsymbol{R_d}^{Bob}).$$

This is not reasonable because the value in the fused information needs does not indicate the portion but the importance degree of information needs. For example, the values of "pork" and "chicken" in the fused information needs are 0.2 and 1.0, but it does not imply that pork and chicken should follow the portion of 1:5

in the resource. In other words, it does not mean that the more similar a resource profile to the fused information needs vector, the higher is the relevance score. Instead, it indicates that the more degree of a resource satisfying the fused information needs, the higher is the relevance score.

If we use keywords match as the measurement, then there will be no difference between the two resources c and d, which is not reasonable either:

$$\phi(\boldsymbol{F}_{Bob}, \boldsymbol{R_c}^{Bob}) = \phi(\boldsymbol{F}_{Bob}, \boldsymbol{R_d}^{Bob}).$$

According to the above analysis, we make the following assumption.

Assumption 3. *Users are more interested in resources which have more tags/terms overlapping with the terms in their fused information needs, and such overlapping tags/terms should have higher values instead of lower values in the resource profiles.*

Based on the above assumption, we propose our revised needs-relevance function as follows:

$$\phi(\boldsymbol{F}_i, \boldsymbol{R}_c^i) = \frac{k}{n} \times \frac{\boldsymbol{F}_i \cdot \boldsymbol{R}_c^i}{\sum_{t(i,x)\in \boldsymbol{F}_i}^{x} v(i,x)}, \tag{13}$$

where k is the number of overlapping terms between social filtering resource profile and fused information needs, n is the total number of tags in the fused information needs. Note that the revised needs-relevance function in (13) is one possible function to solve the problem in Example 3, and fuzzy functions which satisfy Assumption 3 (e.g., $\boldsymbol{F}_i \cdot \boldsymbol{R}_c^i$) can tackle it as well.

Let us revisit Example 3. By adopting (13), we obtain the following result:

$$\phi(\boldsymbol{F}_{Bob}, \boldsymbol{R_c}^{Bob}) = 0.44 < \phi(\boldsymbol{F}_{Bob}, \boldsymbol{R_d}^{Bob}) = 0.91.$$

Therefore, it is more reasonable to use revised needs-relevance function to measure the relevance of resources with respect to the fused information needs. The resources ranked by their relevance scores can finally be returned to the user as the query result.

5 Experiments

5.1 Experiment Setup

5.1.1 Dataset

We collect data from our implemented prototype Folksonomy-based Multimedia Retrieval System (short as FMRS)[34] for personalized recipe search. In the FMRS dataset, there are all 500 recipes in five categories such as Sichuan, Cantonese recipes and so on, 203 users and 7 889 user-generated tags. On average, each user has tagged 16.7 recipes. The tags not only describe various aspects of the recipes but also users' perception on them. In addition, we set the number of user communities as five for the FMRS dataset, and the cluster-based generation is adopted in the experiment.

5.1.2 Metrics

To evaluate our proposed method, we use three different metrics imp (Ranking improvement)[42], P@N (Precision @N)[43] and MRR (Mean reciprocal rank). The first one, *imp*, is to measure how much a personalized ranking list is improved when comparing to the baseline rank. It is defined as follows:

$$\text{imp}(q_i) = \frac{1}{Rank_p(R_{q_i})} - \frac{1}{Rank_b(R_{q_i})}, \tag{14}$$

where q_i is a query, $Rank_p(R_{q_i})$ and $Rank_b(R_{q_i})$ are the rank of target resource R_{q_i} by two different methods to be compared. The overall ranking improvement is the average ranking improvement over all test queries, i.e.,

$$\text{imp} = \sum_{i=1}^{n} \frac{\text{imp}(q_i)}{n}, \tag{15}$$

where n is the number of the test queries. The second metric P@N indicates how accurate a particular personalized search strategy is, and it is calculated by a piecewise function below:

$$\text{P@}N(q_i) = \begin{cases} 1, & \text{if } Rank(R_{q_i}) \leqslant N, \\ 0, & \text{if } Rank(R_{q_i}) > N, \end{cases} \tag{16}$$

where $Rank(R_{q_i})$ is the rank of the target resource, and N is the top N resources in the query result list. Similarly, the overall P@N is calculated as the average P@N by n (the number of queries), as follows:

$$\text{P@}N = \sum_{i=1}^{n} \frac{\text{P@}N(q_i)}{n}. \tag{17}$$

The third metric MRR measures how fast this personalized strategy assists users to find the target resource, i.e.,

$$\text{MRR} = \frac{1}{n} \times \sum_{i=1}^{n} \frac{1}{Rank(R_{q_i})}, \tag{18}$$

where n is the number of queries, $Rank(R_{q_i})$ is the rank of relevant resource in the result list.

5.1.3 Baselines

We use TF-IDF, BM25[7], HYBRID②[6] and NTF[4] as baseline methods to construct collective resource profiles and social filtering resource profiles. We then compare their performance with respect to the above three metrics. Moreover, we compare two search strategies as

②The hybrid of TF-IDF and BM25.

mentioned in Section 4, viz, the cosine similarity ranking method and our proposed revised needs-relevance function (of (13)). In addition, the two different fused methods (linear and switching) are also compared with each other. We summarize the methods to be compared and their abbreviations in Tables 1 and 2, respectively.

Table 1. Abbreviation of Different Strategies

	Strategies		
	Profiling	Ranking	Fusion
C	Collective	Cosine	Linear
S	Social Filtering	Cosine	Linear
S*	Social Filtering	Revised	Linear
S†	Social Filtering	Revised	Switching

Table 2. 16 Methods for Comparison

	BM25	IFIDF	HYBRID	NTF
C	C-BM25	C-TFI	C-HYB	C-NTF
S	S-BM25	S-TFI	S-HYB	S-NTF
S*	S*-BM25	S*-TFI	S*-HYB	S*-NTF
S†	S†-BM25	S†-TFI	S†-HYB	S†-NTF

5.2 Experimental Results

5.2.1 Overall Performance Comparison

The performance comparison in terms of metrics P@N and MRR for the different methods are illustrated in Table 3. From the result, we can see that the methods using social filtering resource profiles (S, S* and S† methods) achieve better performance than those methods by collective resource profiles (C methods). This finding is consistent with our intuition that social filtering resource profiles are more effective to the problem of conflicting tags during the construction of resource profiles (Example 1).

Table 3. Performance Comparison

	P@5	P@10	P@20	MRR
C-BM25	0.092	0.113	0.133	0.094
S-BM25	0.105	0.125	0.149	0.113
S*-BM25	0.112	0.126	0.198	0.169
S†-BM25	0.112	0.128	0.201	0.171
C-TFI	0.105	0.113	0.127	0.112
S-TFI	0.112	0.122	0.169	0.132
S*-TFI	0.141	0.227	0.299	0.146
S†-TFI	0.140	0.231	0.302	0.151
C-HYB	0.105	0.170	0.279	0.118
S-HYB	0.112	0.198	0.342	0.183
S*-HYB	0.126	0.213	0.357	0.192
S†-HYB	0.131	0.218	0.359	0.194
C-NTF	0.192	0.366	0.449	0.198
S-NTF	0.227	0.371	0.459	0.232
S*-NTF	0.270	0.385	0.473	0.237
S†-NTF	**0.271**	**0.387**	**0.478**	**0.240**

Moreover, those S† methods gain improvement on S* methods by varying the fuse method from linear to switching. It shows that the switching fusion method is more reasonable than linear one (Example 2). Besides, the strategies using revised needs-relevance function (S* and S† methods) outperform those with cosine similarity measurement (C and S methods). These results verify that the revised needs-relevance function is more suitable for measuring relevance between user information needs and resource profiles (Example 3).

For each of the baselines (BM25, TFI, HYB or NTF), we further compare its C method to S, S* and S† methods. As shown in Fig.4, no matter what the baseline paradigm is used, it can be further improved by the social filtering resource profiles, the revised needs-relevance function, and the switching fused method. Moreover, we also use the metric imp to each of the methods which achieves the best performance (S† methods) in its paradigm. As shown in Fig.5, when compared with the other three paradigms (BM25, TFI and HYB), NTF has an improvement on the rank from 4.6% to 8.9%. This is consistent with the conclusion in [4], that is, NTF is the most suitable for both user and resource profiles construction.

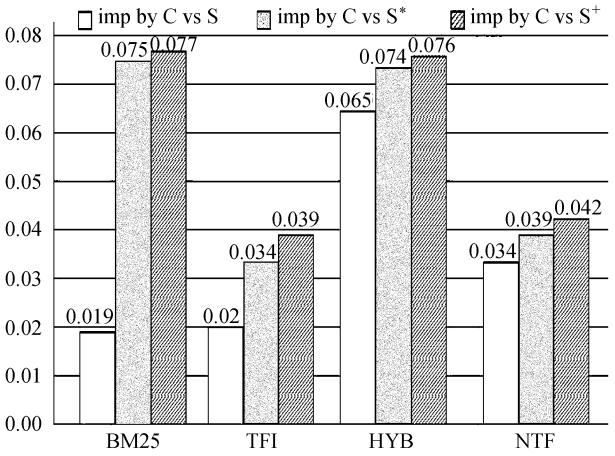

Fig.4. imp metric on different methods.

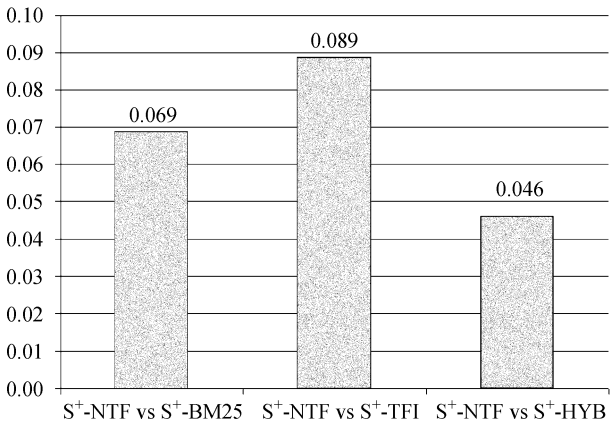

Fig.5. imp metric on different paradigms.

5.2.2 Comparison of Generation Strategies

To compare the three strategies proposed in Section 3 for generating associated relationship between resources and communities, we have invited 10 users who are familiar with cooking to establish the relationships between recipes and communities for our experiment. For heuristic propagation, we use 30% of human established relationships to do the propagation. We select the NTF paradigm in the experiment because it is the most suitable among all paradigms. We examine how these methods impact on the performance, and the result on MRR is shown in Fig.6. According to the result, no matter what the search method (C, S, S* or S†) is adopted, the subjective, heuristic and clustered-based strategies always have the best, secondary, and worst overall performance, respectively. The reason is that the subjective strategy requires full human efforts, while heuristic propagation and cluster-based generation are semi-human and non-human involved strategies. Thus, the subjective strategy is the most precise among the three strategies, and cluster-based is the worst one. But the trade-off is that precision comes from full human efforts, which are costly and not scalable.

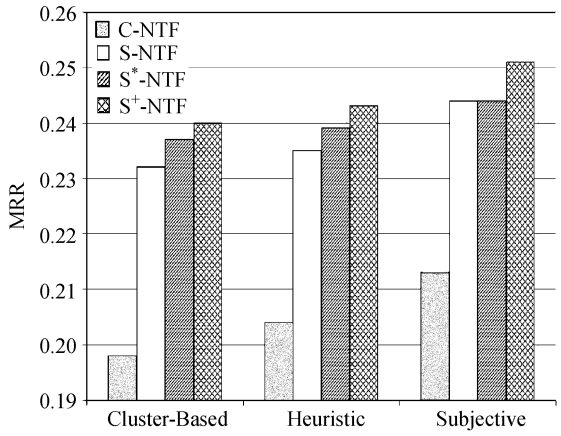

Fig.6. MRR result by different strategies.

For the heuristic propagation, we also measure the speed of propagation by giving different percentages of the initial subset A which is derived from the subjective strategy. Before the experiment, the resources without any tags or which contain self-describing tags are filtered out, and these resources are about 8.5% in our dataset. As shown in Fig.7, we can see that the propagation speed is very fast. Even with only giving the initial percentage of 5%, it will cover all the relationships within 4 iterations. By giving 30%, it only requires two iterations of propagation from the subset to all the relationships.

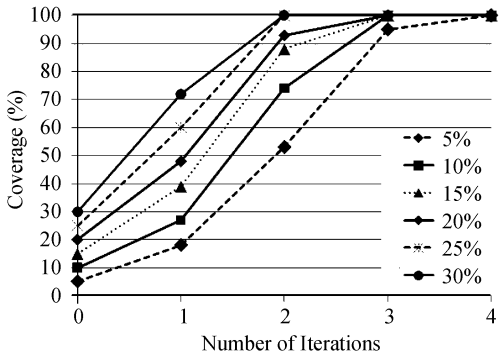

Fig.7. Profanation speed of heuristic strategy.

5.2.3 Community Amount Influence

In addition, we demonstrate the average MRR performances on C, S, S* and S† methods by varying the number of user communities t. As shown in Fig.8, when $t = 1$, the social filtering resource profile (S methods) is equivalent to the collective resource profile (C methods). When t increases (from 1 to 5), the average MRR values for S and S* methods are also increased. This is mainly because resource profiles become more effective when users are grouped into user communities more accurately (i.e., t increases). Moreover, S, S* and S† methods achieve best performance when $t = 5$. Note that this happens to be the number of recipe categories in our FMRS dataset. When t exceeds 5, however, the performance of S, S* and S† methods drops from 5 to 10, since the number of available users for constructing social filtering resource profiles become less and less with the increasing number of user communities. In particular, the sparseness problem negatively influences the performance of S, S* and S† methods when $t > 5$ since FMRS corpus includes 5 categories.

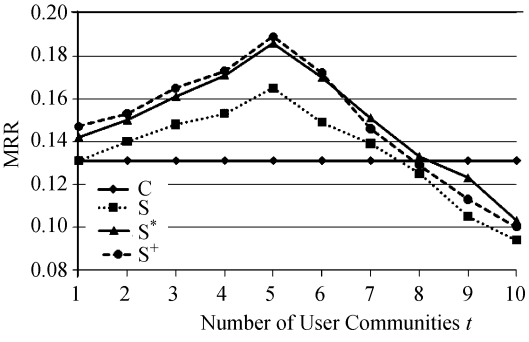

Fig.8. Average MRR as per user community.

5.2.4 Comparison of Cluster-Based Methods

In our last set of experiment, we compare four cluster-based methods: Topic Model[36], Conceptual Space[37], Gaussian Mixture Model[38] and Probability

Model[39] in terms of their average MRR performances on S† by varying the number of user communities t. As illustrated in Fig.9, they all have similar tendency on average MRR performance when varying the number of user communities t, as discussed in the last subsection. They all have the best performance when $t = 5$, and these performances are quite close (from 0.185 to 0.191). This illustrates that our proposed approach has stable (insensitive) and good performance with respect to various cluster-based generation methods when the number of user communities is appropriate. Furthermore, the Conceptual Space[38] has the best performance among these four methods, and the reason can be explained from the following two aspects: 1) not only the connections between tags and resources but also the users are leveraged for the cluster generation process; 2) it focuses on the domain of folksonomy while other three methods focus on other domains such as academic network or e-business.

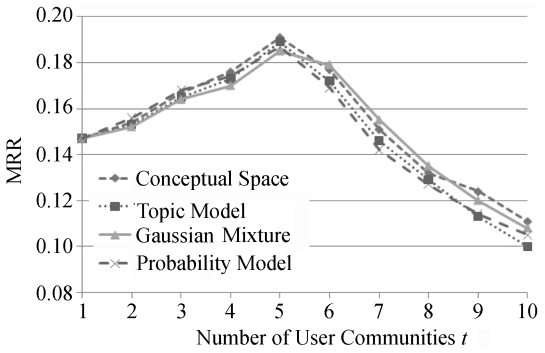

Fig.9. Performance of various cluster-based methods.

6 Conclusions and Future Work

In this paper, we focus on how to address the problem resulted from conflicting user tags for when using existing personalized search methods in folksonomies. We have proposed three strategies to establish resource associated relationships so as to form user communities. By utilizing these communities, a way of constructing resource profiles via social filtering is devised to cope with the conflictive tags problem. Besides, we have devised a personalized search approach by combining a switching fusion method and revised needs-relevance function, so as to optimize resources ranking based on user preferences and a user issued query. Our experimental results show that 1) the social filtering (user community) approach can lead to more precise resource profiles than conventional collective ways do; 2) the switching fusion method is more suitable than straightforward linear one; 3) the revised needs-relevance function is more accurate and reasonable than the intuitive cosine method.

For our upcoming research, we shall continue our study along the following directions. 1) Existing user profiles and resource profiles will be further enhanced by utilizing WordNet③. 2) Rather than using the flat structure of tag:value pair for both user and resource profiles, more sophisticated models such as multi-layered profiling will be adopted for more precise descriptor. 3) Additional information like user query context and relationships between user profiles and contexts will be incorporated to better facilitate personalized search.

References

[1] http://delicious.com/, Sept. 2011.
[2] http://www.movielens.org/, Sept. 2011.
[3] http://www.flickr.com/, Sept. 2011.
[4] Cai Y, Li Q. Personalized search by tag-based user profile and resource profile in collaborative tagging systems. In *Proc. the 19th CIKM*, Oct. 2010, pp.969-978.
[5] Noll M G, Meinel C. Web search personalization via social bookmarking and tagging. In *Proc. the 6th ISWC/ASWC*, Nov. 2007, pp.367-380.
[6] Vallet D, Cantador I, Jose M. Personalizing web search with folksonomy-based user and document profiles. In *Proc. ECIR*, Mar. 2010, pp.420-431.
[7] Xu S, Bao S, Fei B, Su Z, Yu Y. Exploring folksonomy for personalized search. In *Proc. the 31st SIGIR*, Jul. 2008, pp.155-162.
[8] Xie H, Li Q. Resource profiling with social filtering for personalized search in collaborative tagging systems. In *Proc. SWSM/SIGIR*, Jul. 2011, pp.35-43.
[9] Golder S A, Huberman A. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 2006, 32(2): 198-208.
[10] Bischoff K, Firan C S, Nejdl W, Paiu R. Can all tags be used for search? In *Proc. the 17th CIKM*, Oct. 2008, pp.193-202.
[11] Gupta M, Li R, Yin Z, Han J. Survey on social tagging techniques. *SIGKDD Explorations Newsletter*, 2010, 12(1): 58-72.
[12] Bao S, Xue G, Wu X, Yu Y, Fei B, Su Z. Optimizing web search using social annotations. In *Proc. the 16th WWW*, May 2007, pp.501-510.
[13] Michlmayr E, Cayzer S. Learning user profiles from tagging data and leveraging them for personal(ized) information access. In *Proc. the 16th TMSCO/WWW*, May 2007.
[14] Markines B, Cattuto C, Menczer F, Benz D, Hotho A, Gerd S. Evaluating similarity measures for emergent semantics of social tagging. In *Proc. the 18th WWW*, Apr. 2009, pp.641-650.
[15] Almeida R B, Almeida V A F. A community-aware search engine. In *Proc. the 13th WWW*, May 2004, pp.413-421.
[16] Park L A F, Ramamohanarao K. Mining web multi-resolution community-based popularity for information retrieval. In *Proc. the 16th CIKM*, Nov. 2007, pp.545-554.
[17] Smyth B. A community-based approach to personalizing web search. *Computer*, 2007, 40(8): 42-50.
[18] Deng H, King I, Lyu M R. Enhancing expertise retrieval using community-aware strategies. In *Proc. the 18th CIKM*, Nov. 2009, pp.1733-1736.

③http://wordnet.princeton.edu/.

[19] Teevan J, Morris M R, Bush S. Discovering and using groups to improve personalized search. In *Proc. the 2nd WSDM*, Feb. 2009, pp.15-24.
[20] Teevan J, Dumais S T, Horvitz E. Potential for personalization. *ACM Transaction on Computer-Human Interaction*, 2010, 17(1): Article No. 4.
[21] Liu N, Yang Q. EigenRank: A ranking-oriented approach to collaborative filtering. In *Proc. the 31st SIGIR*, Jul. 2008, pp.83-90.
[22] Ma Z, Pant G, Sheng O R L. Interest-based personalized search. *ACM Transaction on Information Systems*, 2007, 25(1): Article No.5.
[23] Carmel D, Zwerdling N, Guy I, Ofek-Koifman S, Har'el N, Ronen I, Uziel E, Yogev S, Chernov S. Personalized social search based on the user's social network. In *Proc. the 18th CIKM*, Nov. 2009, pp.1227-1236.
[24] Song W, Zhang Y, Liu T, Li S. Bridging topic modeling and personalized search. In *Proc. the 23rd COLING*, Aug. 2010, pp.1167-1175.
[25] Su Z, Yang Q, Zhang H, Xu X, Hu Y. Correlation-based document clustering using web logs. In *Proc. the 34th HICSS*, Jan. 2001, p.5022.
[26] Qiu F, Cho J. Automatic identification of user interest for personalized search. In *Proc. the 15th WWW*, May 2006, pp.727-736.
[27] Dou Z, Song R, Yuan X, Wen J R. Are click-through data adequate for learning web search rankings? In *Proc. the 17th CIKM*, Oct. 2008, pp.73-82.
[28] Wang Q, Jin H. Exploring online social activities for adaptive search personalization. In *Proc. the 19th CIKM*, Oct. 2010, pp.999-1008.
[29] Leung K W T, Fung H Y, Lee D L. Constructing concept relation network and its application to personalized web search. In *Proc. EDBT/ICDT*, Mar. 2011, pp.413-424.
[30] Lee J W, Kim H J, Lee S. Applying taxonomic knowledge to bayesian belief network for personalized search. In *Proc. SAC*, Mar. 2010, pp.1796-1801.
[31] Sendhilkumar S, Geetha T V. Personalized ontology for web search personalization. In *Proc. the 1st COMPUTE*, Jan. 2008, Article No.18.
[32] Xue G R, Han J, Yu Y, Yang Q. User language model for collaborative personalized search. *ACM Transaction on Information Systems*, 2009, 27(2): Article No. 11.
[33] Dou Z, Song R, Wen J R. A large-scale evaluation and analysis of personalized search strategies. In *Proc. the 16th WWW*, May 2007, pp.581-590.
[34] Cai Y, Li Q, Xie H, Yu L. Personalized resource search by tag-based user profile and resource profile. In *Proc. the 11th WISE*, Dec. 2010, pp.510-523.
[35] Fischer G. External and shareable artifacts as opportunities for social creativity in communities of interest. In *Proc. the 5th CCMCD*, Dec. 2001, pp.67-89.
[36] Tang J, Jin R, Zhang J. A topic modeling approach and its integration into the random walk framework for academic search. In *Proc. the 8th ICDM*, Dec. 2008, pp.1055-1060.
[37] Wu X, Zhang L, Yu Y. Exploring social annotations for the semantic web. In *Proc. the 15th WWW*, May 2006, pp.417-426.
[38] Zhang H, Giles C, Foley H C, Yen J. Probabilistic community discovery using hierarchical latent Gaussian mixture model. In *Proc. the 22nd AAAI*, Jul. 2007, pp.663-668.
[39] Zhou D, Manavoglu E, Li J, Giles C L, Zha H. Probabilistic models for discovering e-communities. In *Proc. the 15th WWW*, May 2006, pp.173-182.
[40] Blei D, Ng A, Jordan M. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 2003, 3(1): 993-1022.
[41] Papoulis A. Probability, Random Variables, and Stochastic Processes. New York: McGraw-hill, 1965.
[42] Shepitsen A, Gemmell J, Mobasher B, Burke R. Personalized recommendation in social tagging systems using hierarchical clustering. In *Proc. RecSys*, Oct. 2008, pp.259-266.
[43] White R W, Bailey P, Chen L. Predicting user interests from contextual information. In *Proc. SIGIR*, Jul. 2009, pp.363-370.

Hao-Ran Xie received the B.Eng. degree in software engineering from Beijing University of Technology, China, and the M.Sc. degree in computer science from the City University of Hong Kong, Kowloon. He is currently a Ph.D. candidate in the Department of Computer Science, the City University of Hong Kong. His research interests include user modeling, personalized search, recommendation system and multimedia retrieval and management.

Qing Li received the B.Eng. degree from Hunan University, China, and the M.Sc. and Ph.D. degrees from the University of Southern California, Los Angeles, USA, all in computer science. He is currently a professor at the City University of Hong Kong, Kowloon, a visiting professor at Zhejiang University, Hangzhou, China, a guest professor at the University of Science and Technology of China, Hefei, and an adjunct professor at Hunan University. His research interests include database modeling, Web services, multimedia retrieval and management, and e-learning systems. He has authored over 200 papers in technical journals and international conferences. He is actively involved in the research community by serving as a journal reviewer, program committee chair/co-chair, and organizer/co-organizer of several international conferences. Prof. Li serves as the chairman of the Hong Kong Web Society, a councilor of the Database Society of China Computer Federation, and a Steering Committee member of the International WISE Society.

Yi Cai received the B.Eng. degree in polymer engineering and M.Sc. degree in computer science from Sichuan University, China, and Ph.D. degree in computer science from the Chinese University of Hong Kong. He is currently an assistant professor of School of Software Engineering at the South China University of Technology, Guangzhou, China. His research interests are recommendation system, personalized search, semantic Web, cognitive modeling, and data mining.