# Personalized Search by Tag-based User Profile and Resource Profile in Collaborative Tagging Systems

Yi Cai
Department of Computer Science
City University of Hong Kong
Hong Kong, China
yicai3@cityu.edu.hk

Qing Li
Department of Computer Science
City University of Hong Kong
Hong Kong, China
itqli@cityu.edu.hk

## ABSTRACT

With the increase of resource-sharing web sites such as YouTube[1] and Flickr [2], personalized search becomes more important and challenging, as users demand higher retrieval quality. To achieve this goal, personalized search needs to take users' personalized profiles and information needs into consideration. Collaborative tagging (also known as folksonomy [11]) systems allow users to annotate resources with their own tags, which provide a simple but powerful way for organizing, retrieving and sharing different types of social resources. In this paper, we examine the limitations of previous tag-based personalized search. To handle these limitations, we propose a new method to model user profiles and resource profiles in a collaborative tagging environment. A novel search method using such users' and resources' profiles is proposed to facilitate the desired personalization in resource search. We implement a prototype system named as FMRS. Experiments using FMRS data set and MovieLens data set show that our proposed method outperforms baseline methods.

## Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information filtering*

## General Terms

Algorithms, Experimentation, Human Factors

## Keywords

Collaborative Tagging, Personalized Search, User Profiling

## 1. INTRODUCTION

The advent of resource-sharing web sites like Flickr and YouTube allows users to share multimedia resources with each other, and has brought huge amount of resources to the web. Resource search becomes more important and challenging, as users demand higher retrieval quality. Current resource search methods mainly depend on the relevant match of the query and resource descriptions. Although different users input the same query terms, they may have different information needs corresponding to their own preferences. Thus, it is necessary to implement personalized resource retrieval based on users' profiles so as to obtain more pertinent and useful search results.

Currently, collaborative tagging systems become more and more popular and many social resource sites support tagging mechanism. For example, bookmarks on Del.icio.us [3] may be tagged in terms of topics interesting to the user; in Flickr, users can upload and annotate their own photos; in Last.fm [4], users can annotate their favorite songs. The resources and the tags posted by Web users to these systems are supposed to be highly dependent on their interests, and the tags given by users provide rich information for building more accurate and specific user profiles [1]. Further, the tags given by different users to a resource are useful to describe the resource. It provides a collaborative viewpoint on resource description and such a description is considered as more meaningful and acceptable from users' perspectives [6] [2].

According to these characteristics of collaborative tagging systems, researchers consider that constructing users' profiles and resources' profiles from collaborative tags is instrumental for personalized resource search. Some research works such as [9] [13] are conducted to construct user and resource profiles from tags in collaborative tagging systems, and also some personalized search methods are proposed based on tags such as [17] [23] [22]. However, some limitations exist in the current tag-based personalized search methods, including the following:

- In previous works, the weights of tags in a user profile or resource profile are based on TF (term frequency), or TF-IDF (term frequency-inverse document frequency),[5] or BM25. By using absolute term frequency as tag weights in user profiles, it will enlarge the weights of active users who frequently annotate resources and decrease the weights of other users who rarely annotate resources. For TF-IDF, IDF is used to reflect how well a tag can distinguish a user (or resource) from others only, and it is not necessary nor useful in indicating a user's preferences on tags or how a resource is relevant to tags. As TF and IDF are variables of the BM25 function, the latter suffers the same limitations by nature. Section 2.4 gives a more detailed discussion.

- In current works on tag-based personalized search, the match

---

[1]http://www.youtube.com
[2]http://www.flickr.com

---

[3]http://delicious.com
[4]http://www.last.fm
[5]IDF refers to IUF (inverse user frequency) for user profile construction and to IRF (inverse resource frequency) for resource profile construction in this paper.

of a user and a resource is based on the similarity between the user's profile and the resource's profile. However, the resources that are relevant to a user's interests actually need to possess the user's favorite tags as many as possible (instead of as similar as possible). Usually, the weight of a tag in a resource profile is the degree of how relevant the resource is to the tag, while the weight of a tag in a user profile is the favor degree of the user on the tag. It does not mean better if the weight of a tag in a user profile is more similar to the weight of the corresponding tag in a resource profile. For instance, it is not the best match between a resource and a user when all the same tags' weights in both of the user profile and the resource profile are small. The reason is that a small weight of a tag $x$ in a user profile means that the user is interested in $x$ to a small extent instead of not interested in it, while a small weight of a tag $y$ in a resource profile may mean that the resource is not so relevant to $y$. Therefore, although the similarity value between $x$ and $y$ is high, the tagged resource may not be relevant to the user's interests. In our viewpoint, the problem of how a resource satisfies a user's interest requirement is more of a fuzzy satisfaction problem, instead of a similarity measure problem. More discussions and illustrations are given in section 2.5.

In this paper, we reveal and elaborate on the limitations of current works on tag-based personalized search. To handle these limitations, we propose a new tag-based model for constructing user and resource profiles. Based on the tag-based user profiles and resource profiles, we further propose a novel personalized search framework in a collaborative tagging system. The new features of the proposed framework and the contributions of our work are as follows.

- For user profile construction, we use a normalized term frequency (NTF) to indicate the preference degree of a user on a tag. For a given tag $x$, NTF is the possibility or proportion of a user using $x$, and we regard NTF more appropriate to reflect how much the user is interested in $x$. The resource profiles are similarly constructed based on NTF also.

- Instead of keyword match or similarity measurement as in previous works, relevance measurement between a resource and a user query, termed as *query relevance*, is treated in our framework as a fuzzy satisfaction problem of a user's query requirements. If a resource can satisfy as many as possible the user query requirements (in terms of all the query terms), then the resource will be more (content) relevant. Besides, based on an observation on user query behavior, we present a query relevance function which takes the number of matching tags between the query and a resource into consideration.

- Similar to the above query relevance measurement, relevance measurement between a resource and a user's interest requirement, termed as *user (interest) relevance*, is considered as a fuzzy satisfaction problem. If a resource can satisfy a user's interests (i.e., tags in user profiles) as many as possible, then the resource will be regarded as more relevant.

- To illustrate and validate our proposed approach, we have implemented a prototype personalized recipe retrieval system named as Folksonomy-based Multimedia Retrieval System (FMRS). In FRMS, a recipe can be introduced to people through multimedia resources such as texts of recipe introduction, videos demonstrating the cooking procedure of the recipe, or photos of the dishes corresponding to the recipes. Experimental studies are conducted upon FMRS to evaluate the effectiveness of our approach.

- In addition to FMRS, we also conduct experiments with real data from MovieLens [6]. The experiment results show that our method outperforms the state-of-the-arts methods in personalized resource search.

To the best of our knowledge, this is the first effort to model query relevance measurement and the user (interest) relevance measurement as fuzzy satisfaction problems.

The rest of this paper is organized as follows. Section 2 introduces the background and related work, where we also reveal and discuss the limitations of the existing works. In section 3, we model user and resource profiles in a collaborative tagging system. We propose a personalized search method in section 4 using tag-based user and resource profiles. In section 5, experiments are conducted to compare our proposed method with previous methods on FMRS and MovieLens data sets. Section 6 concludes the paper and introduces potential future works.

## 2. BACKGROUND AND RELATED WORKS

In this section, we first survey some existing works on collaborative tagging and personalized search. Next we examine and discuss the limitations of these works in terms of user and resource profiling, as well as user (interest) relevance measurement.

### 2.1 Collaborative Tagging

Existing research on collaborative tagging can be divided into two types. The first type focuses on investigating and analyzing the features of tags. Golder and Huberman [12] analyzed the tag usage patterns, user activities and annotated resources in collaborative tagging systems. Bischoff et al. [5] did a deep survey to systems like Del.icio.us, and Last.fm[7], to discover useful tags for information access.

The second type mainly attempts to explore social annotations and link structures in folksonomy for various applications. Bao et al. [4] proposed two algorithms, named as SocialSimRank and SocialPageRank respectively, to explore the latent semantics of tags to optimize web search. A recent work done by Markines et al. [15] did evaluations of various similarity measures on semantics of social tags. In [16], the author proposed that tag based profiles can be represented by naive approach, co-occurrence approach, adaptive approach, etc. However, there are drawbacks of this profiling method, either due to the lack of user input which makes discovering co-occurred tags difficult, or the difference existing in personally preferred vocabulary which leads to problems like polysemous and synonymy tags, and thus influences the precision of learned profiles. Chirita et al. [3] proposed a method which automatically generates personalized tags for Web pages. Au Yeung et al. [14] presented their analysis on personal data in folksonomies and investigated how accurately user profiles can be generated from those data. Through the tag-based profiles, personalized search [23] in collaborative tagging systems became possible and popular, as it can facilitate users greatly to find interesting resources.

### 2.2 Personalized Search

The current strategies of personalized search fall into two categories [19]:

- One is query expansion such as [7], which refers to modifying the original query either by expanding it with other terms, or assigning different weights to the terms in the query.

- Another category is result processing, primarily re-ranking, which adapts the search results to a particular user's preference. Most re-ranking strategies attempt to construct a user profile from the user's historical behavior, and use the profile to filter out resources unmatching with his/her interests. Chirita et al. [8] modeled both user profiles and resources as topic vectors from ODP [8] hierarchy, thus the matching between user interest and content can be measured by their vector distance. A personalized PageRank algorithm was proposed in [18], which was a modification to the global PageRank on Web, and the search results were personalized based on the hyperlink structure. Besides learning user profiles based on their own browsing histories, Sugiyama et al. [21] also explored social information to refine search results with the help of like-minded neighbors. Dou et al. [10] did comparisons between various personalization approaches, like click based, profile based, long-term based, short-term based, etc., and proposed an evaluation framework for the strategies.

## 2.3 Personalized Search in Collaborative Tagging Systems

With the recent development of collaborative tagging systems, some works are proposed on personalized search in the collaborative tagging environment. Noll and Meinel [17] propose a simple yet effective approach to explore user's and resource's related tags based on term frequency, and re-rank the non-personalized search results based on these related tags. Xu et al. [23] propose topic-based personalized search in folksonomy, in which the personalized search is conducted by ranking the resources based on not only term similarity matching but also topic similarity matching. In their work, instead of using term frequency, term frequency-inverse document frequency (TF-IDF) and BM25 are used to construct user and resource profiles. As a sequel to [23], Vallet et al. [22] use different techniques to measure the user-resource similarities and compare the effect of these techniques.

Although there are several works to handle personalized search with tag-based user and item profiles, they have some limitations. In next subsections, we examine and discuss these limitations.

## 2.4 Limitations of User and Resource Profile Construction

In current works on tag-based personalized search, user profiles and resource profiles are constructed based on the Vector Space Model (VSM) or BM25 ranking model. The weight of each tag in a user profile is the degree to which the user is interested in the tag. In addition, the weight of each tag $x$ in a resource profile is the degree to which the resource is relevant to the tag.

In the Vector Space Model, e.g., the queries, resources and users are all mapped to tag vectors in a universal term space [23]. Generally, the weight of each tag in a user profile (vector) is calculated by term frequency (TF), or term frequency-inverse user frequency (TF-IUF) as follows:

$$w = tfiuf = tf \times \log \frac{N}{n_i}$$

where $tf$ denotes the term frequency, $N$ denotes the total number of users and $n_x$ denotes the number of users using tag $x$. Similarly, the weight of each tag in a resource profile is calculated by TF, or term frequency-inverse resource frequency (TF-IRF).

For BM25 ranking model, the weights of tags in a user profile is

measured as follows:

$$w = bm25_{u_m}(t_x) = iuf(t_x) \cdot \frac{tf \cdot (k_1 + 1)}{tf + k_1 \cdot (1 - b + b \cdot \frac{ul}{avgul})}$$

where $k_1$ and $b$ are two parameters often set to the standard values of 2 and 0.75, $ul$ is the length of the user profile vector and $avgul$ is the average length of all user profiles. Similarly, the weights of tags in a resource profile are measured as follows:

$$bm25_{r_m}(t_x) = irf(t_x) \cdot \frac{tf \cdot (k_1 + 1)}{tf + k_1 \cdot (1 - b + b \cdot \frac{dl}{avgdl})}$$

where $dl$ is the length of the resource profile vector and $avgdl$ is the average length of all resource profiles.

However, using TF only or TF-IUF or BM25 values as weights of tags can be rather insufficient in indicating how much a user is interested in the tags, as the following discussion is going to reveal.

### 2.4.1 Limitation of TF

For using TF as tag weights in user profiling, if some active users frequently annotate resources while other users rarely annotate resources, then the tag weights of the more active users will be much higher than those of less active ones. Let us consider the following example.

**Example 1.** *For two users Bob and Alice, if Bob is an active user who frequently gives tags to resources and he has annotated 300 resources, while Alice is a less active user who only gives tags to resources occasionally and she has annotated 30 resources only. Then suppose their user profiles are calculated by TF only, as shown below:*

$$\overrightarrow{U}_{Bob} = (spicy : 135, chicken : 198, \cdots, sweet : 164)$$

$$\overrightarrow{U}_{Alice} = (spicy : 25, chicken : 28, \cdots, sweet : 23)$$

*Based on these two profiles, we cannot conclude that Bob is much more interested in "spicy", "chicken" and "sweet" dishes than Alice is. Actually, Alice can be also quite (or even most) interested in these three tags, because we can find that she has used these three tags more than 22 times among her 30 times resource annotations. The probability of Alice likes to use these three tags should be even greater than that of Bob.*

Therefore, TF is itself not sufficient in reflecting user preferences on tags. Similarly, TF is not good enough to indicate how much a tag is relevant to a resource.

### 2.4.2 Limitation of TF-IUF and TF-IRF

To use TF-IUF as tag weights in user profiling, on the other hand, we note that IUF is actually unsuitable for indicating a user's preferences on tags. The weights are only used in reflecting how well the tags can distinguish a user from others, and they are not useful in reflecting a user's preferences. Sometimes, there can be side-effect. Let us consider the following example.

**Example 2.** *There are 1000 users, and suppose for a user Tom and the information of his tagging is as shown in table 1, where $n_x$ is the number of users who use tag $x$. If we adopt TF-IUF to measure the tag's weights for Tom's profile, then the IUF value of "wasabi" will be higher than that of other tags because $n_{wasabi}$ is smaller than that of other tags. Hence, we will have the following relationships among the tag's weights:*

$$w_{wasabi} > w_{spicy} > w_{chicken} > w_{sweet}$$

*However, this result is neither intuitive nor reasonable. Tom rarely uses the tag "wasabi" for annotation but often uses the other three*

| | spicy | chicken | $\cdots$ | sweet | wasabi |
|---|---|---|---|---|---|
| $TF$ | 38 | 41 | $\cdots$ | 44 | 6 |
| $n_x$ | 500 | 820 | $\cdots$ | 850 | 1 |

**Table 1: The tagging information of Tom**

*tags, as he is actually more interested in the other three tags as opposed to "wasabi".*

The IUF can help distinguish a user from other users but can not help reflect how much a user is interested in a particular tag. But our purpose of constructing a user profile is to reflect the user's preferences on tags so as to facilitate personalized search, instead of trying to distinguish one user from the others. Thus, we consider it as unnecessary to combine IUF with TF for indicating user preferences in our user profiling approach.

Similarly, TF-IRF is also unsuitable for reflecting how a tag is relevant to a resource.

### 2.4.3 Limitation of BM25

For BM25, since TF and IUF(IRF) are variables of the BM25 function, it follows naturally that the limitations of TF and TF-IUF (TF-IRF) are also applicable to BM25 when it is used to calculate weights of tags for constructing user and resource profiles.

## 2.5 Limitation of User Interest Relevance Measurement

For personalized search, personalization is achieved by measuring resources' relevance to a user's interests. Thus, it depends on matching a user's profile with resources' profiles in order to find out the resources most relevant to that user's preferences. In current works on tag-based personalized search, the match of a user profile and a resource profile is based on similarity calculation. The more similar they are, the resource matches the user's interests better. A commonly used similarity function is cosine similarity which has been adopted in many current works such as [23] [22].

Assume that measuring a resource relevance to a user's interests is indeed a similarity measurement problem, then a user profile should be corresponding to an "ideal" resource (the most favor resource) for the user, and the weights of tags in that user profile are most similar to those of tags in the ideal resource profile. However, actually, the weights of a user profile indicate the preference degrees of the user on his favor tags. In a user's profile, all tags with weights greater than zero are somewhat favored tags by the user. The larger the weight is, the more interested the user will be in the tag. On the other hand, a small value of a tag weight does not indicate that the user is not interested in the tag; instead, it means that the preference degree of the user on the tag is relatively small when compared with those tags with higher weights. Obviously, in all current works, weights of tags in a user profile are not corresponding to those of tags in the profile of the resource most favored by the user. For a resource profile, the weight of a tag in it indicates the degree of relevance between the resource and the tag. The higher the degree, the more relevant the resource is to the tag.

In matching a user profile with resource profiles, the resources which are regarded as relevant to a user's interests need to possess the user's favored tags as many as possible. It does not mean "better" (i.e., the user may not be more interested in the resource) if the weight of a tag in the user profile is more similar to the weight of the corresponding tag in a resource profile. Let us take a look at the following example.

**Example 3.** *Assume a user $i$ has his profile as follows:*

$$\overrightarrow{U}_i = (spicy : 0.8, chicken : 0.5, \cdots, sweet : 0.25)$$

*According to the profile of user $i$, he likes spicy taste very much, and likes chicken in general and sweet taste a little bit (the weight 0.25 does not mean that he does not like "sweet"). In other words, this profile means that user $i$ cares about spicy taste more than chicken, and far more than sweet taste in his preference. Now suppose there are two resources c and d, their profiles are as follows:*

$$\overrightarrow{R}_c = (spicy : 0.95, chicken : 0.95, sweet : 0.85)$$

$$\overrightarrow{R}_d = (spicy : 0.9, chicken : 0.1, sweet : 0.01)$$

*Resource c is very relevant to the tags "spicy", "chicken" and "sweet". For resource d, although it is also very relevant to "spicy", it is less relevant to "chicken" and "sweet". If we use a similarity function (e.g., cosine similarity) to measure how relevant the two resources are for user $i$, we can obtain the following result:*

$$Sim(\overrightarrow{U}_i, \overrightarrow{R}_c) < Sim(\overrightarrow{U}_i, \overrightarrow{R}_d)$$

*This result is however against intuition. For user $i$, he is interested in all of the three features with different degrees. For resource c, it is very relevant to all of these features. For resource d, it is only relevant to "spicy" and not so relevant to "chicken" and "sweet". Therefore, intuitively, user $i$ should be more interested in c than d.*

Based on the above analysis, we can conclude that similarity function is not suitable for matching the relevance between a user profile and a resource profile.

## 3. USER AND RESOURCE PROFILING

One important issue in personalized search is how to represent or describe a user's profile and a resource's profile, i.e., using what kinds of techniques to measure the a degree a user is interested in a feature (tag) and the degree a resource is relevant to a feature. In this section, different from current mainstream techniques on modeling user and resource profiles, we propose to model user and resource profiles based on normalized term frequency (NTF).

### 3.1 Modeling User Profiles

In a folksonomy-based system, a user can annotate resources by tags and the tags can reflect this user's preference to certain extent. As mentioned earlier, tag frequency (TF) measure and inverse user frequency (IUF) are not suitable/adequate in indicating a user's preference degrees on tags for annotation. Based on our observation on user behaviors, we make the following assumption:

**Assumption 1.** *For user $i$ and two given tags $x$ and $y$, if user $i$ uses tag $x$ more frequently than $y$ to annotate different resources, we assume that user $i$ is more interested in tag (feature) $x$ than $y$.*

According to assumption 1, how much user $i$ is interested in tag $x$ depends on the possibility or proportion of his using tag $x$ to annotate resources among all his annotations. Based on assumption 1, we can represent a user profile as follows:

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

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

where $t_{i,x}$ is a tag that is interested by user $i$, $n$ is the total number of tags that user $i$ may be interested in, $v_{i,x}$ is the preference degree of user $i$ on tag $t_{i,x}$ and , intuitively, can be obtained as follows:

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

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 $v_{i,x}$, the more preferred (favorable) is user $i$ on tag $x$. Note that $v_{i,x}$ is actually the normalized term frequency for tag $x$ used by user $i$.

Intuitively, if a user uses a tag more frequently for annotation, then it means the user is more interested in the tag. Thus, term frequency can be used to indicate a user's interests on a tag to some extent. However, there are some users who are more active than others, and these active users annotate resources by tags more frequently than inactive users. If we only take absolute term frequency into consideration on tag weighting for each user, higher weights will be given to the tags in the active users' profiles than those in an inactive user's profile. Thus, it is necessary to normalize the term frequency of each tag for each user so as to avoid the bias on active users. Adopting the possibility or proportion of a user's using a specific tag to reflect the preference of the user on that tag is more appropriate. For example, for two users $i$ and $j$, user $i$ has used tag $x$ for 9 times among all his 10 times annotations, while user $j$ has used $x$ for 10 times among all his 100 times annotations. According to assumption 1, user $i$ is more interested in tag $x$ than user $j$ because the proportion of his using tag $x$ for annotation (9 out of 10) is greater than that of user $j$ (10 out of 100). This is intuitively more reasonable than adopting TF or TF-IUF which only considers the absolute number of times a user has used a particular tag.

Let us revisit the examples 1 and 2 in section 2. For example 1, if we use NTF values as weights of tags in user profiles, we can obtain the profiles of Bob and Alice as follows.

$$\overrightarrow{U}_{Bob} = (spicy : 0.45, chicken : 0.66, \cdots, sweet : 0.5467)$$

$$\overrightarrow{U}_{Alice} = (spicy : 0.833, chicken : 0.933, \cdots, sweet : 0.767)$$

Obviously, such a result alleviates the bias on active users. For example 2, we can obtain the following profile of Tom (suppose Tom has annotated 100 resources) by using our method:

$$(spicy : 0.38, chicken : 0.41, \cdots, sweet : 0.44, wasabi : 0.06)$$

Thus, the relationships of weights of these tags in Tom's profile are as follows.

$$w_{sweet} > w_{chicken} > w_{spicy} > w_{wasabi}$$

According to assumption 1, Tom uses tags "sweet", "chicken" and "spicy" more frequently and rarely uses the tag "wasabi" for annotation, thus he should be more interested in the other three tags instead of "wasabi". This result is more reasonable and objective to reflect user interests.

## 3.2 Modeling Resource Profiles

Similar to user profile modeling, we make an assumption below on how a tag represents a resource.

**Assumption 2.** *For a resource d and two tags x and y, if the number of users who use tag x to annotate resource d is greater than that of users who use y to annotate d (i.e., when compared with y, tag x is more popularly used to annotate d), we assume that more users regard tag (feature) x as being more salient than tag y for resource d.*

According to assumption 2, how well tag $x$ is used to describe resource $d$ is dependent on the possibility or proportion of users' using tag $x$ to annotate resource $d$ among all users who annotate $d$. Based on assumption 2, we can represent a resource profile as follows:

**Definition 2.** A **resource profile** for a resource $c$, denoted by $\overrightarrow{R_c}$ by is a vector of tag:value pairs:

$$\overrightarrow{R_c} = (t_{c,1} : w_{c,1}, t_{c,2} : w_{c,2}, \cdots, t_{c,n} : w_{c,n})$$

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 the 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 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 $c$. Note that $w_{c,x}$ is actually the normalized term frequency (NTF) of tag $x$ used to annotate resource $c$, and it can indicate the common popular degree of using tag $x$ to represent resource $c$ among the users who have annotated $c$.

We regard it as more suitable to use the normalized term frequency to reflect the relevance of a tag to a resource. Intuitively, if more users choose to use a particular tag to describe a specific resource, then it means the tag is a more salient feature of that resource and more representative for that resource. As an example, for a tag $x$, suppose there are 10 users who annotate resource $c$ and 9 users among them use tag $x$ to annotate $c$, while there are 100 users who annotate resource $d$ and 10 users among them use tat $x$ to annotate $d$. According to assumption 2, tag $x$ is more representative for resource $c$ than $d$ because the proportion of users who choose to use tag $x$ for annotating $c$ is greater than that for annotating $d$.

## 4. PERSONALIZED SEARCH

In a personalized search system, users have different information needs (usually represented in the form of user input queries) and different personal interests. Due to the reason that there may be many resources that can satisfy a user query, the search results should also satisfy and match the user's personal interests, by letting the user find out what he or she actually wants quickly (e.g., by putting the expected result on the top of the search result). Thus, personalized search is to find out the information that not only satisfy a user's basic information need and but also best match his or her personal interests. This process can be split into two subprocesses. One is query relevance measurement which is to find out to what extent resources satisfy a user's basic information need. Another is user interest relevance measurement, which is to re-rank the result by advancing the resources in the result that best satisfy the user's personal interests.

In our personalization framework, we first compute the content relevance between a query and resources, then measure how much the resources match with a user's interest, and aggregate the two relevance scores to get the final rankings as the results.

## 4.1 Query Relevance Measurement

In our work, we assume a user query is usually in the form of a vector of terms.

**Definition 3.** A **query** issued by user $i$ denoted by $\overrightarrow{q_i}$ is a vector of terms as follows:

$$\overrightarrow{q_i} = (t_{i,1}^q, t_{i,2}^q, \cdots, t_{i,m}^q)$$

where $t_{i,x}^q$ is a term, and $m$ is the total number of terms in the query. For example, a user may issue a query of "chicken" if he wants to find a chicken dish, or "spicy fish" to search for dishes made of fish with spicy flavor and so on.

The objective of traditional search is to rank resources based on their relevance to a given query. It is achieved by matching the query with resource profiles. The relevant score of a resource for

a query can be formally measured by a **query relevance function** as follows:

$$\gamma : Q \times R \rightarrow [0, 1]$$

where $R$ is the set of resources, $Q$ is the set of queries. The result of $\gamma$ function is a relevant score of a resource to a query, and the higher relevant score, the more relevant is the resource to the query.

In a personalized search system, users' information needs are in the form of the user input queries. A user issued query includes basic content requirements of the user on resources. In our approach, the query relevance measurement is considered as a fuzzy satisfaction problem as per a user query requirements. More specifically, we regard a user query as fuzzy requirements of the user on resources' content and each term in the query is a fuzzy requirement for the relevant resources. If a resource can satisfy the user requirements as many as possible, then the resource will be more (content) relevant. Based on this consideration, we present some axioms for the function $\gamma$ to observe.

**Axiom 1.** For a resource $c$ and a query $\overrightarrow{q}_i$, if there does not exist a tag $x$ where $t^q_{i,x} \in \overrightarrow{R}_c$, then $\gamma(\overrightarrow{q}_i, \overrightarrow{R}_c) = 0$.

**Axiom 2.** For a resource $c$ and a query $\overrightarrow{q}_i$, if $\forall x, t^q_{i,x} \in \overrightarrow{R}_c$ and $w_{c,x} = 1$, then $\gamma(\overrightarrow{q}_i, \overrightarrow{R}_c) = 1$.

**Axiom 3.** For two resources $c$ and $d$ and a query $\overrightarrow{q}_i$, if $\exists x, t_{c,x} \in \overrightarrow{q}_i, t_{d,x} \in \overrightarrow{q}_i, w_{c,x} > w_{d,x}$, and $\forall y, y \neq x, t_{c,y} \in \overrightarrow{q}_i, t_{d,y} \in \overrightarrow{q}_i, w_{c,y} = w_{d,y}$, then $\gamma(\overrightarrow{q}_i, \overrightarrow{R}_c) > \gamma(\overrightarrow{q}_i, \overrightarrow{R}_d)$.

Axioms 1 and 2 specify the boundary cases of measuring the content relevance degree of a resource for a query. If there is no intersection between a query and a resource profile, then the resource does not satisfy any requirement of the user (i.e., does not match any term in the query issued by the user and is irrelevant to the query), hence $\gamma(\overrightarrow{q}_i, \overrightarrow{R}_c) = 0$. On the other hand, if all terms in the query $\overrightarrow{q}_i$ are definitely salient in the resource profile $\overrightarrow{R}_c$, then $\gamma(\overrightarrow{q}_i, \overrightarrow{R}_c) = 1$. Axiom 3 specifies the essence of the query relevance function. If a resource profile satisfies (possesses) more query terms with higher degrees, then the resource will be more relevant to the query.

The following is a possible function which satisfies axioms 1-3.

$$\gamma(\overrightarrow{q}_i, \overrightarrow{R}_c) = \frac{\sum w_{c,x}}{m}, t_{c,x} \in \overrightarrow{q}_i \qquad (3)$$

where $m$ is the total number of terms in the query $\overrightarrow{q}_i$.

In equation 3, the query relevance measurement does not take the number of matched keywords (co-occurring keywords) between a query and a resource profile into consideration. However, we have the following observation on user query behaviors, which indicates that the number of matched keywords has an effect on the query relevance measurement.

**Observation 1.** *The relationship among all the terms in a query is conjunctive, meaning that when a user issues a query, he wants to find out relevant resources which possesses all the query terms instead of just a part of them.*

Thus, according to observation 1, a resource satisfying all the query terms is better (more pertinent) than that of the resource only satisfying the query terms partially. The following example illustrates this observation.

**Example 4.** *User Bob wants to search some recipes about spicy chicken and he inputs the query terms "spicy" and "chicken". Suppose there are three recipes c, d and e whose resource profiles are shown as follows.*

$$\overrightarrow{R}_c = (spicy : 0.3, chicken : 0.5, \cdots)$$

$$\overrightarrow{R}_d = (spicy : 0.2, chicken : 0.3, \cdots)$$

$$\overrightarrow{R}_e = (spicy : 0.95, chicken : 0, \cdots)$$

*If we only consider the linear combination of all the weights of them according to equation 3, then the query relevance will have the following relationship:*

$$\gamma(\overrightarrow{q}_i, \overrightarrow{R}_e) > \gamma(\overrightarrow{q}_i, \overrightarrow{R}_c) > \gamma(\overrightarrow{q}_i, \overrightarrow{R}_d)$$

*However, recipe e only satisfies one query terms and should not be more relevant than c and d. If we use traditional keyword match and do not consider the weights of tags in the resource profiles, the result will be as follows:*

$$\gamma(\overrightarrow{q}_i, \overrightarrow{R}_e) < \gamma(\overrightarrow{q}_i, \overrightarrow{R}_c) = \gamma(\overrightarrow{q}_i, \overrightarrow{R}_d)$$

*This result can not identify that resource c is better than d.*

To rectify the problem, we propose a **revised $\gamma$ function** as follows:

$$\gamma(\overrightarrow{q}_i, \overrightarrow{R}_c) = \frac{\sum w_{c,x}}{m} \cdot (\frac{k}{m})^\alpha, t_{c,x} \in \overrightarrow{q}_i \qquad (4)$$

where $k$ is the number of the terms satisfied by resource $c$ in query $\overrightarrow{q}_i$, $m$ is the total number of terms in the query and $\alpha$ is a parameter used to adjust the effect of the number of relevant tags in a resource profile for a query.

It is easy to verify that equation 4 also observes axioms 1 to 3. Let us revisit example 4. If we use equation 4 to measure the query relevance of resources $c$, $d$ and $e$, we can obtain the following result:

$$\gamma(\overrightarrow{q}_i, \overrightarrow{R}_c) > \gamma(\overrightarrow{q}_i, \overrightarrow{R}_d) > \gamma(\overrightarrow{q}_i, \overrightarrow{R}_e)$$

This result is intuitively more reasonable than the one obtained by using equation 3.

## 4.2 User Interest Relevance Measurement

Resources with high content-relevance to a query may not always be accepted by the user who issues the query, for the resources may deviate from his/her personal favor. The objective of user interest relevance measurement is to identify how a user would be interested in a resource. Such a process can be formally defined by a function as follows.

$$\theta : U \times R \rightarrow [0, 1]$$

where $R$ is the set of resources, $U$ is the set of users. The result of $\theta$ function is a relevant score of a resource to a user's personal interests, and the higher value of $\theta$, the more relevant the resource is to the user's personal interests.

As mentioned above, the value $v_{i,x}$ of each tag $x$ in the profile of user $i$ is the preference (interestedness) degree of $i$ on tag $x$. If a user is more interested in a tag, he will also be more likely interested in the resources which are relevant to the tag. In other words, a user demands (i.e., expects to find) resources which are more relevant to the tags that he is more interested in. If a resource can satisfy a user's personal interests as many as possible, then the resource will be more relevant to him. From this point of view, similar to query relevance measurement, the user relevance measurement is treated as a fuzzy satisfaction problem on user interest requirement. Tags in a user profile are considered as interest requirements of the user in personalized search. The higher weight value of a tag, the more strict requirement of the user on the tag. Based on this consideration, in our method, the matching degree of the profile of user $i$ and the profile of resource $c$ is measured by a

**user interest relevance function** $\theta(\overrightarrow{U}_i, \overrightarrow{R}_c)$. The following are the axioms for this function to observe.

**Axiom 4.** For a resource $c$ and a user $i$, if $\forall x, v_{i,x} > 0, w_{c,x} = 0$, then $\theta(\overrightarrow{U}_i, \overrightarrow{R}_c) = 0$.

**Axiom 5.** For a resource $c$ and a user $i$, if $\forall x, v_{i,x} > 0, w_{c,x} = 1$, then $\theta(\overrightarrow{U}_i, \overrightarrow{R}_c) = 1$.

**Axiom 6.** For user $i$, let $c$ and $d$ be two resources, if $\exists x, v_{i,x} > 0, w_{c,x} > w_{d,x}$, and $\forall y, y \neq x, w_{c,y} = w_{d,y}$, then $\theta(\overrightarrow{U}_i, \overrightarrow{R}_c) > \gamma(\overrightarrow{U}_i, \overrightarrow{R}_d)$.

**Axiom 7.** For two users $i$ and $j$, let $c$ be a resource, if $\exists x, w_{c,x} > 0, v_{i,x} < v_{j,x}$, and $\forall y, y \neq x, w_{c,y} > 0, v_{i,x} = v_{j,x}$, then $\theta(\overrightarrow{U}_i, \overrightarrow{R}_c) > \gamma(\overrightarrow{U}_j, \overrightarrow{R}_c)$.

Axioms 4 and 5 specify the boundary cases of measuring the user interest relevance degree of a resource to a user. If a resource does not possess any tags in a user's profile, then the resource does not satisfy (match) any requirement of the user interests, hence $\theta(\overrightarrow{U}_i, \overrightarrow{R}_c) = 0$. If a resource satisfies all the terms in a user profile $\overrightarrow{U}_i$ definitely, then $\theta(\overrightarrow{q}_i, \overrightarrow{R}_c) = 1$. Axioms 6 and 7 specify, respectively, the influence of the tag weights on function $\theta$ in user profiles and resource profiles.

The following is a possible function which satisfies axioms 4-7:

$$\theta(\overrightarrow{U}_i, \overrightarrow{R}_c) = \frac{\sum l_x}{m} \tag{5}$$

where $m$ is the total number of terms in the query, and

$$l_x = \begin{cases} w_{c,x} + (1 - v_{i,x})(1 - w_{c,x}) & 1 > w_{c,x} > 0, v_{i,x} > 0 \\ 1 & w_{c,x} = 1, v_{i,x} > 0 \\ 0 & w_{c,x} = 0, v_{i,x} > 0 \end{cases} \tag{6}$$

Equation 5 considers all tags as of the same importance, i.e., users will treat all the tags at the same interesting level. However, we have the following observation on user behaviors.

**Observation 2.** *Users are more interested in tags which carry higher preference values, and they will give more weights to tags in which they are more interested.*

Based on observation 2, we revise equation 5 and propose a **weighted user interest relevance function** as follows:

$$\theta(\overrightarrow{U}_i, \overrightarrow{R}_j) = \frac{\sum l_x \cdot v_{i,x}}{m} \tag{7}$$

It is easy to verify that equation 7 also observes axioms 4 to 7.

Let us revisit the example 3 in section 2. Based on the equations 6 and 7 in our method, we can obtain the following result:

$$\theta(\overrightarrow{U}_i, \overrightarrow{R}_c) > \theta(\overrightarrow{U}_i, \overrightarrow{R}_d)$$

This result is intuitive and reasonable because resource $c$ satisfies all requirements of the profile of user $i$, while resource $d$ only satisfies the requirement of the tag "spicy".

## 4.3 Personalized Ranking

The final goal of personalized search is to get resources that match both the query requirements and user's personal interest. Based on the query relevance value obtained by $\gamma$ function (e.g., equation 4) and the user interest relevance value obtained by $\theta$ (e.g., equation 7) function, we aggregate both of them into a final personalized relevant score so as to obtain the final resource ranking for a particular user query. The final personalized relevance score between a resource and a query issued by user $i$ is calculated as follows:

$$RScore(\overrightarrow{q}_i, \overrightarrow{U}_i, \overrightarrow{R}_c) = \frac{\gamma(\overrightarrow{q}_i, \overrightarrow{R}_c) + \theta(\overrightarrow{U}_i, \overrightarrow{R}_c)}{2} \tag{8}$$

where $\overrightarrow{q}_i$ is a query, $\overrightarrow{R}_c$ is a resource profile, $\overrightarrow{U}_i$ is a user profile, all of which are in the form of term vectors. The higher the value of $RScore(\overrightarrow{q}_i, \overrightarrow{U}_i, \overrightarrow{R}_c)$ is, the higher position resource $c$ should be in the result list of query $\overrightarrow{q}_i$ issued by user $i$. The reason is that resource $c$ is highly relevant to not only the query but also the user's interests. We do not adopt the multiplication aggregation here because there can be some resources which are relevant to query (i.e., $\gamma(\overrightarrow{q}_i, \overrightarrow{R}_c) > 0$) but not relevant to user's interest at all (i.e., $\theta(\overrightarrow{U}_i, \overrightarrow{R}_c) = 0$), if we use multiplication to aggregate $\gamma(\overrightarrow{q}_i, \overrightarrow{R}_c)$ and $\theta(\overrightarrow{U}_i, \overrightarrow{R}_c)$, then $RScore(\overrightarrow{q}_i, \overrightarrow{U}_i, \overrightarrow{R}_c)$ is zero. Thus, those resources which are relevant to a query to some extent but not relevant to a user's interest at all will have the same ranking score (i.e., zero) as resources which are irrelevant to both the query and the user's interest.

## 5. EXPERIMENT

In this section, we conduct experiments to compare our approach with other current personalized search methods in collaborative tagging systems. As part of our work, we have implemented a prototype of personalized recipe retrieval system named as FRMS. It allows users to tag and collect their preferred recipes in a collaborative manner. In this section, we report experiments conducted upon FRMS and MovieLens data to compare our method with three state-of-the-art methods.

### 5.1 Data Set

In our experiments, two different data sets are used. One is a small data set upon FRMS, and another is a larger data set of MovieLens.

In FRMS, there is a database of 500 recipes and 203 users. Besides, there are 7889 user-resource-tag tuples and each user on average annotated 12 recipes. These tags mainly describe recipes from the aspects of ingredient, taste, nutrition and so on. On average, each user annotates 16.7 resources, and each resource has 15.778 tags as its annotations.

To further evaluate our method, we also use a larger data set of MovieLens. This data set consists of 95,580 user-movie-tag tuples annotated by 71,567 users on 10,681 movies. Each tag is typically a single word, or a short phrase. The meaning and purpose of a particular tag are determined by each user.

For each of the two data sets, we randomly split the data set into two parts, 80 percent of tags as the training set and 20 percent of tags as the test set. We use the data in the training set to construct user profiles and resource profiles. Based on the constructed profiles, we use the data in the test set as input queries to test the efficiency of the personalized search methods. Similar to all the other collaborative tagging systems, there are lots of polysemy and synonym words in the collected tags, and a simple preprocessing was conducted to reduce ambiguity.

### 5.2 Evaluation Metrics

To evaluate the efficiency of our method, three metrics are employed here for the evaluation.

The first one is $imp$ as presented in [20]. This is a common evaluation metric to measure how a personalization strategy improves the ranking of the target resources of a user in the result list by comparing to baseline methods. It is defined as:

$$imp(q_i) = \frac{1}{r_p} - \frac{1}{r_b} \tag{9}$$

where $q_i$ is an issued query, $r_b$ is the rank of target resource by a baseline search approach, and $r_p$ is the rank of the same resource
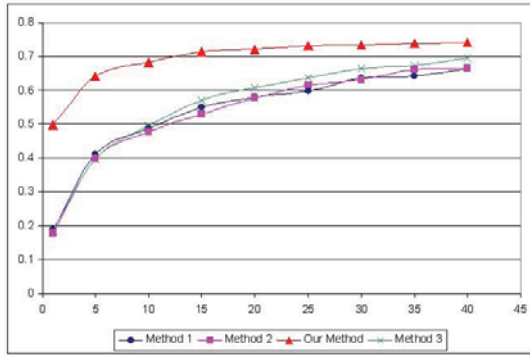
**Figure 1: Comparison of our method and baseline methods on *hitrate* with different $HR@n$ using FRMS data set**



**Figure 2: Comparison of our method and baseline methods on *imp* using FRMS data set**

returned by our personalized search. The overall *ranking improvement* is calculated as "average query *imp*" for all the queries in the test data, as follows:

$$imp = \frac{\sum_{i=1}^{m} imp(q_i)}{m}$$

where $m$ is the number of queries. A larger value of $imp$ indicates a greater improvement of the ranking for target resource by the personalization approach.

The second metric we use is *hitrate* (HR) [13], which is used to measure how often user interested resources are in the recommendation or personalized search result list. It is defined as:

$$HR(u_i) = \frac{|T_{u_i} \bigcap X_{u_i}|}{|T_{u_i}|} \qquad (10)$$

where $T_{u_i}$ denotes the resources relevant to (interested by) user $i$ in the test set, and $X_{u_i}$ is the result set of top-N returned resources. The overall *hitrate* of the top-N results is computed as "average $HR(u_i)$" for all the users in the test data set, as follows:

$$HR = \frac{\sum_{i=1}^{n} HR(u_i)}{n} \qquad (11)$$

where $n$ is the number of users. The larger the average *hitrate* is, the more precise the personalized search model is.

The third metric is *Mean reciprocal rank(MRR)*, which is a statistic for evaluating a ranking to a query. The reciprocal rank of a query result is the multiplicative inverse of the rank of the first correct answer. The mean reciprocal rank is the average of the reciprocal ranks of results for a query. It is defined as follows:

$$MRR = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{rank_i} \qquad (12)$$

where $m$ is the number of queries, $rank_i$ is the position of the correct answer (relevant resource) in the result ranking for the query $i$. The larger the average *MRR* is, the faster and easier for the user to find out the resources he or she wants.

## 5.3 Baseline Methods

To evaluate the effectiveness of our method, we compare our approach with three state-of-the-art personalized search methods in collaborative tagging systems. The first one (Method 1) is the method presented in [23], with the weights of tags in user profiles and resource profiles being based on TF-IDF values. The second one (Method 2) also comes from [23] with the weights of tags
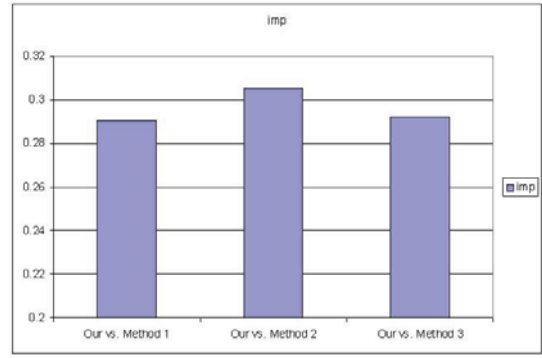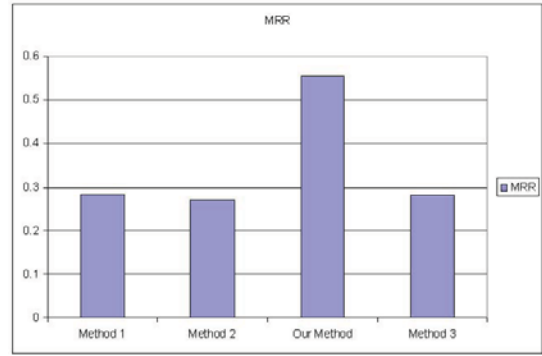


**Figure 3: Comparison of our method and baseline methods on *MRR* using FRMS data set**

in user profiles and resource profiles being based on BM25 values. The third method (Method 3) is a state-of-the-art personalized search method from [22], in which the weights of tags in user profiles and resource profiles are an aggregation of BM25 values and TF-IDF values. These three methods resemble the current mainstream techniques for handling personalized search in collaborative tagging systems, and they use different methods to model user and resource profiles.

## 5.4 Experiment Results

### 5.4.1 On FRMS Data Set

We first compare our method with the baseline methods by using FRMS data. Figure 1 shows the comparison of our method and the compared methods with different $n$ (i.e., size of result sets) on *hitrate* metric for FRMS data set, where $hitrate@n$ means that the size of the returned result set is $n$.

From figure 1, we can see that our method outperforms all the three compared methods for all values of $n$ (i.e., from $n = 1$ to $n = 40$).[9] The higher value of $n$ is, the higher value of *hitrate* all methods get. We can find that when $n = 40$ (i.e., 40 resources are returned in the result list), our approach can achieve a *hitrate* at 0.7417, which means most (74.19%) of users can find the resources they want. Even when $n = 1$, our method still can obtain

---

[9]We limit $n$ from 1 to 40 here because most users would like to find out what they want quickly by just browsing a few returned results and should not expect them to review too much returned results.
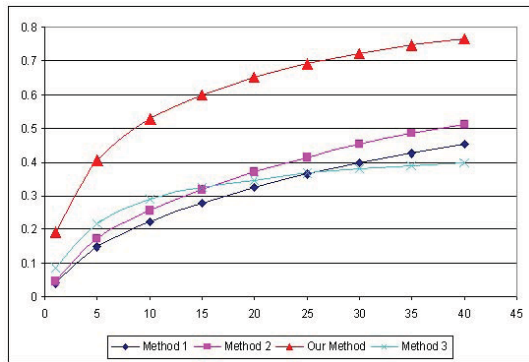
**Figure 4: Comparison of our method and baseline methods on** *hitrate* **with different** $HR@n$ **using MovieLens data set**



**Figure 5: Comparison of our method and baseline methods on** *imp* **using MovieLens data set**
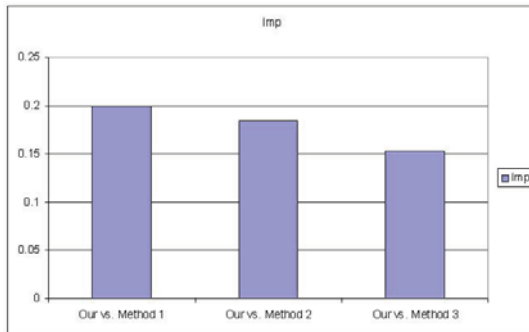


**Figure 6: Comparison of our method and baseline methods on** *MRR* **using MovieLens data set**

a *hitrate* at 0.4981. In other words, for most queries, users can find out the exact resource they want by just browsing top-40 returned resources. With the increase of $n$, the increase of *hitrate* of all methods become slower and slower after $n$ is 10 or more. For our method, the *hitrate* can reach 0.6826 while $n = 10$, whereas it is 0.7417 when $n = 40$. The reason is that the data set of FRMS we use is not of a big size (500 resources only) and the search space is relatively small, thus we can get most of resources we want by just returning a small set of resources. For the same reason, *hitrate* of all the methods become convergent as the $n$ increases.

Figure 2 shows the comparison of our method and the other three methods on *imp* metric using FRMS data set. We can find that our method outperforms method 1 by 29.03 %, method 2 by 30.54% and method 3 by 29.21 % on *imp*. Figure 3 shows the comparison of our method and the other methods on *MRR* using FRMS data set. According to Figure 3, our method obtains the highest MRR value at 0.5551 while the other compared methods are only around 0.275. Our method outperforms the compared methods by at least 94.17% (the best compared method is method 1 whose MRR is 0.2833). According to these results, we can conclude that for FRMS data set, our method can push the user favorite resources to a more front position in the result ranking list enabling users to find what they want easier and faster.

### 5.4.2 *On MovieLens Data Set*

To further evaluate our method, we conduct experiments on Movie-Lens data set which is a much bigger data set. Figure 4 shows
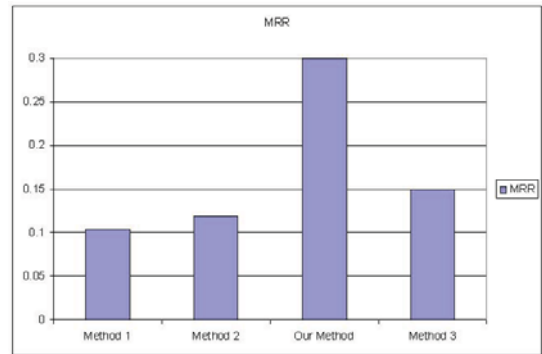
the comparison of our method and the compared methods with different $n$ on *hitrate* metric using MovieLens data set. Similar to the result of FRMS data set, our method outperforms all the three compared methods for all values of $n$. Again the *hitrate* value of all methods increases as $n$ increases. Different from the result of FRMS, our method's *hitrate* value is only at 0.1923 for Movie-Lens data set when $n = 1$. Similarly, the other compared methods obtain worse results than those for FRMS data set. We think the reason is that for FRMS data set which is related to recipe domain, tags mainly describe recipes from the aspects of ingredient, taste, nutrition and so on. Thus, the tag space and resource space are relatively small and it is easier to find out the expected resources for users. For MovieLens data, there are many genres of movies and the tags given by users are from wider aspects. The tag space and resource space are much greater than that of FRMS, so the precision of search is not as good as that of FRMS.

Figure 5 shows the comparison of our method and the other three methods on *imp* metric for MovieLens data set. According to Figure 5, our method outperforms method 1 by 19.92 %, method 2 by 18.44% and method 3 by 15.25 % on *imp*. Figure 6 shows the comparison of our method and the other methods on *MRR* for MovieLens data set. Our method obtains the highest MRR value at 0.2999 while the other compared methods are only around 0.1037 to 0.1497. Our method outperforms the compared methods by at least 100.3% (the best compared method is method 3 whose MRR is 0.1497).

Based on Figures 1 to 6, we can conclude that our method outperforms all the compared methods on all adopted metrics for both FRMS and MovieLens data sets, due to two main differences between our method and previous methods. One is how to measure weights of tags in user and resource profile, and another is how to measure query relevance and user interest relevance. The experiment results demonstrate the advantage and effectiveness of our method on data sets of different scale.

## 6. CONCLUSION AND FUTURE WORKS

In this paper, we focus on exploring personalized resource search by tag-based user profiles and resource profiles. We have proposed a new method to model user profiles and resource profiles from a new perspective in collaborative tagging environment. We have then proposed a search method using such user and resource profiles, thereby facilitating the much desired personalization in resource search. There are two main differences between our method and previous methods. One is that we adopt normalized term fre-

quency (NTF) as weights of tags in user and resource profiles. The other difference is to treat query relevance measurement and user interest relevance measurement as fuzzy requirement satisfaction problems. To the best of our knowledge, this is the first effort on this direction. A prototype of personalized recipe retrieval system named as Folksonomy-based Multimedia Retrieval system (FMRS) in a collaborative tagging environment is implemented. Experiments in FMRS and MovieLens data show that our proposed method outperforms previous methods and can improve the personalized resources search effectively.

There are several potential future extensions to our work. Currently, we use a simple aggregation function to combine the query relevance and user interest relevance, and we will explore more on the aggregation function to study the effect of aggregation on the search result. Besides, in our current work, we focus on the relationship among users, tags and resources, and construct user and resource profiles based on tags. In collaborative tagging systems, users can form different communities. We foresee it to be quite useful to take the community information into consideration in modeling user and resource profiles so as to enhance the personalized search in collaborative tagging systems.

# 7. ACKNOWLEDGEMENT

# 8. REFERENCES

[1] H. Al-Khalifa and H. Davis. Measuring the semantic value of folksonomies. In *Innovations in Information Technology, 2006*, pages 1–5, November 2006.

[2] H. S. Al-Khalifa and H. C. Davis. Exploring the value of folksonomies for creating semantic metadata. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 3(1):13–39, March 2007.

[3] P. Alex, R. Chirita, S. Costache, S. H, and W. Nejdl. P-tag: Large scale automatic generation of personalized annotation tags for the web. In *Proc. of the 16th Intl. World Wide Web Conf*, pages 8–12, 2007.

[4] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su. Optimizing web search using social annotations. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 501–510. ACM, 2007.

[5] K. Bischoff, C. S. Firan, W. Nejdl, and R. Paiu. Can all tags be used for search? In *CIKM '08*, pages 193–202. ACM, 2008.

[6] C. Cattuto, D. Benz, A. Hotho, and G. Stumme. Semantic grounding of tag relatedness in social bookmarking systems. In *ISWC '08: Proceedings of the 7th International Conference on The Semantic Web*, pages 615–631, Berlin, Heidelberg, 2008. Springer-Verlag.

[7] P. A. Chirita, C. S. Firan, and W. Nejdl. Personalized query expansion for the web. In *SIGIR '07*, pages 7–14. ACM, 2007.

[8] P. A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschütter. Using odp metadata to personalize search. In *SIGIR '05*, pages 178–185, New York, NY, USA, 2005. ACM.

[9] J. Diederich and T. Iofciu. Finding communities of practice from user profiles based on folksonomies. In *Proceedings of TEL-CoPs'06*, 2006.

[10] Z. Dou, R. Song, and J.-R. Wen. A large-scale evaluation and analysis of personalized search strategies. In *WWW '07*, pages 581–590. ACM, 2007.

[11] S. A. Golder and B. A. Huberman. The structure of collaborative tagging systems. *CoRR*, abs/cs/0508082, 2005.

[12] S. A. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. *J. Inf. Sci.*, 32(2):198–208, 2006.

[13] H.-I. J. J. J. G. Kim, H.N. User preference modeling from positive contents for personalized recommendation. *Corruble, V., Takeda, M., Suzuki, E.(eds.) DS 2007. LNCS (LNAI)*, 4755:116Í C126, 2007.

[14] C. man Au Yeung, N. Gibbins, and N. Shadbolt. A study of user profile generation from folksonomies. In P. Dolog, M. Krötzsch, S. Schaffert, and D. Vrandecic, editors, *SWKM*, volume 356 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2008.

[15] B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and G. Stumme. Evaluating similarity measures for emergent semantics of social tagging. In *WWW '09*, pages 641–650. ACM, 2009.

[16] E. Michlmayr. Learning user profiles from tagging data and leveraging them for personal(ized) information access. In *Proceedings of the Workshop on Tagging and Metadata for Social Information Organization, WWW2007*, 2007.

[17] M. G. Noll and C. Meinel. Web search personalization via social bookmarking and tagging. In *ISWC'07/ASWC'07: Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference*, pages 367–380, Berlin, Heidelberg, 2007. Springer-Verlag.

[18] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web, 1999.

[19] J. Pitkow, H. Schütze, T. Cass, R. Cooley, D. Turnbull, A. Edmonds, E. Adar, and T. Breuel. Personalized search. *Commun. ACM*, 45(9):50–55, 2002.

[20] A. Shepitsen, J. Gemmell, B. Mobasher, and R. Burke. Personalized recommendation in social tagging systems using hierarchical clustering. In *RecSys '08*, pages 259–266. ACM, 2008.

[21] K. Sugiyama, K. Hatano, and M. Yoshikawa. Adaptive web search based on user profile constructed without any effort from users. In *WWW '04*, pages 675–684. ACM, 2004.

[22] D. Vallet, I. Cantador, and J. M. Jose. Personalizing web search with folksonomy-based user and document profiles. In *Advances in Information Retrieval, 32nd European Conference on IR Research, ECIR 2010, Milton Keynes, UK, March 28-31, 2010. Proceedings*, pages 420–431, 2010.

[23] S. Xu, S. Bao, B. Fei, Z. Su, and Y. Yu. Exploring folksonomy for personalized search. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 155–162, New York, NY, USA, 2008. ACM.