Analyzing Cross-System User Modeling on the Social Web

Fabian Abel, Samur Araújo, Qi Gao, and Geert-Jan Houben

Web Information Systems, Delft University of Technology
{f.abel,s.f.cardosodearaujo,q.gao,g.j.p.m.houben}@tudelft.nl

Abstract. In this article, we analyze tag-based user profiles, which result from social tagging activities in Social Web systems and particularly in Flickr, Twitter and Delicious. We investigate the characteristics of tag-based user profiles within these systems, examine to what extent tag-based profiles of individual users overlap between the systems and identify significant benefits of cross-system user modeling by means of aggregating the different profiles of a same user.

We present a set of cross-system user modeling strategies and evaluate their performance in generating valuable profiles in the context of tag and resource recommendations in Flickr, Twitter and Delicious. Our evaluation shows that the cross-system user modeling strategies outperform other strategies significantly and have tremendous impact on the recommendation quality in cold-start settings where systems have sparse information about their users.

1 Introduction

Social tagging comes in different flavors. Social bookmarking systems like Delicious allow people to tag their bookmarks, in photo sharing platforms such as Flickr users annotate images, and in micro-blogging systems like Twitter people can assign so-called hash tags to their posts. In the last decade we saw a variety of research efforts in the field of social tagging systems ranging from the analysis of social tagging structures [9,7], via information retrieval in folksonomy systems [4,11], to personalization [16,17]. However, most studies have been conducted in the context of particular systems, for example, analyzing the tagging behavior within Delicious [7] or computing tag recommendations in Flickr based on the tagging characteristics within the Flickr system [17]. Our research complements these studies and investigates tagging behavior and user modeling across system boundaries to support engineering of Social Web systems that aim for personalization.

We therefore identified users, who have an account at Flickr, Twitter *and* Delicious, and crawled more than 2 million tagging activities which were performed by these users in the three systems. Based on this dataset, we study the tagging behavior of the same user in different systems. We furthermore present and analyze various cross-system user modeling strategies. For example, we show that it is possible to generate tag-based profiles based on Twitter activities for improving tag and bookmark recommendations in Delicious.

1.1 Personomies and Tag-Based Profiles

The emerging structure that evolves from social tagging is called folksonomy. A folksonomy is basically a set of tag assignments, user-tag-resource bindings attached with some timestamp that indicates when a tag assignment was performed. We base our research on the folksonomy model $\mathbb{F} = (U, T, R, Y)$ proposed by Hotho et. al [11], where U, T, R and Y refer to the sets of users, tags, resources and tag assignments respectively. We also model Twitter posts by means of tag assignments, i.e. we consider each post (*tweet*) as a resource and hash tags such as "#icwe2011", which appear in the post, are—without the "#" symbol—treated as tags. Given the folksonomy model $\mathbb{F}$, the user-specific part of a folksonomy, the *personomy*, can be defined as follows (cf. [11]).

Definition 1 (Personomy). *The personomy* $\mathbb{P}_u = (T_u, R_u, Y_u)$ *of a given user* $u \in U$ *is the restriction of* $\mathbb{F}$ *to* u, *where* T_u *and* R_u *are finite sets of* tags *and* resources *respectively that are referenced from tag assignments* Y_u *performed by the user* u.

While the personomy specifies the tag assignments that were actually performed by a specific user, the *tag-based profile* $P(u)$ is an abstraction of the user that represents the user as a set of of weighted tags (cf. [8,15]).

Definition 2 (Tag-based profile). *The* tag-based profile *of a user* u *is a set of weighted tags where the weight of a tag* t *is computed by a certain strategy* w *with respect to the given user* u.

$$P(u) = \{(t, w(u, t)) | t \in T_{source}, u \in U\} \tag{1}$$

$w(u, t)$ *is the weight that is associated with tag* t *for a given user* u. T_{source} *is the source set of tags from which tags are extracted for the tag-based profile* $P(u)$.

The weights associated with the tags in a tag-based profile $P(u)$ do not necessarily correspond to the tag assignments in the user's personomy $\mathbb{P}_u$. For example, $P(u)$ may also specify the weight for a tag t_i that does neither occur in the personomy $\mathbb{P}_u$ nor in the folksonomy $\mathbb{F}$, i.e. where $t_i \notin T_u \wedge t_i \notin T$. With $P(u)@k$ we describe the subset of a tag-based profile $P(u)$ that contains the k tag-weight pairs that have the highest weights. $\bar{P}(u)$ denotes the tag-based profile for which the weights are normalized so that the sum of all weights in $P(u)$ is equal to 1, and with $|P(u)|$ we refer to the number of distinct tags contained in the tag-based profile.

1.2 Problem Definition

Today, individual users may be active in various social tagging systems so that tag-based profiles about the same user are distributed across different systems. How do tag-based profiles of the same user differ from system to system? To what extent do these profiles overlap? Can cross-system user modeling support the construction of tag-based profiles and how does it impact the performance of recommender systems in cold-start situations where user profiles are sparse? In this paper we answer these questions and tackle the following research challenge.

Definition 3 (User modeling challenge). *Given a user u, the user modeling strategies have to construct a tag-based profile P(u) so that the performance of tag and resource recommendations in cold-start situations is maximized.*

Following related work (e.g. [17]) we define the recommendation tasks as ranking problems:

Definition 4 (Cold-start Recommendation challenge). *Given a tag-based user profile $P(u)$, the personomy of the user $\mathbb{P}_{u,target} = (T_u, R_u, Y_u)$ and a set of tags T_{target} and a set of resources R_{target}, which are not explicitly connected to u ($T_u \cap T_{target} = \emptyset$ and $R_u \cap R_{target} = \emptyset$), the challenge of the recommendation strategies is to rank those tags $t \in T_{target}$ and resources $r \in R_{target}$ so that tags/resources that are most relevant to the user u appear at the very top of the ranking.*

For both tag and resource recommendations, we do not aim to optimize the recommender algorithm itself, but we identify those user modeling strategies that support the recommender algorithms best.

1.3   Related Work

Today, several research efforts aim to support re-use of user profile data on the Social Web. With standardization of exchange languages (e.g. FOAF[1]) and APIs (e.g. OpenSocial[2]), standardized authentication protocols such as OpenID[3], solutions for identifying users across systems [6] and research on generic user modeling [13,10,3], cross-system user modeling becomes tangible. In previous work we analyzed the nature of social networking profiles on the Social Web and introduced a service for aggregating tag-based profiles [2]. Szomszor et al. presented an approach to merge user's tag clouds from two different social networking websites to generate richer user interest profiles [18], but did not investigate the impact of the generated profiles on personalization.

Different from personalization in social tagging systems that targets single systems [16,17], cross-system personalization makes the investments in personalizing a system transferable to other systems. Mehta et al. showed that cross-system personalization makes recommender systems more robust against spam and cold start problems [14]. However, Mehta et al. could not test their approaches on Social Web data where individual user interactions are performed across different systems and domains, but experiments have been conducted on user data, which originated from one system and was split to simulate different systems. In contrast to that, in this paper we evaluate cross-system user modeling and its impact on cold-start recommendations on real world datasets from three typical Social Web systems: Flickr, Twitter and Delicious.

---

[1] `http://xmlns.com/foaf/spec/`
[2] `http://code.google.com/apis/opensocial/`
[3] `http://openid.net/`

2 User Modeling Strategies

The cross-system user modeling strategies that we discuss in this paper consist of the following building blocks: (1) source of user data, (2) semantic enrichment, (3) weighting scheme.

2.1 Source of User Data

In order to construct a user profile and adapt functionality to the individual users, systems require information about their users [12]. The data that is exploited to create user profiles (cf. T_{source} in Definition 2) might come from different sources. In our evaluation (see Section 4) we will therefore compare the following sources of information.

Target Personomy Tags $T_{\mathbb{P}_{target}}$**.** The traditional approach to user modeling is to exploit the user-specific activities observed in the target system that aims for personalization. In a folksonomy system one would thus exploit the personomy of the user $\mathbb{P}_u$ (see Definition 1) that is inferred from the folksonomy $\mathbb{F}_{target}$ of the target system.

Target Folksonomy Tags $T_{\mathbb{F}_{target}}$**.** If the personomy of the user $\mathbb{P}_u$ is rather sparse or even empty, one has to find other sources of information that are applicable to generate a user profile. Therefore, we define another baseline strategy that considers all activities performed in the target system as if they would have been performed by the given user. Hence, by considering the complete folksonomy $\mathbb{F}_{target}$ one obtains some sort of *average profile* which promotes these tags that are popular for the complete folksonomy.

Foreign Personomy Tags $T_{\mathbb{P}_{foreign}}$**.** If the user's personomy in the target system is sparse or empty then another strategy is to utilize the personomy from another system. For example, if a user u has not annotated any resource in Delicious yet, this strategy harnesses the hash tags the user assigned to her Twitter posts, i.e. u's personomy from Twitter.

2.2 Semantic Enrichment

Semantics of tag-based profiles are not well defined. For example, tag-based recommender systems may encounter situations where relevant resources are discarded because tags assigned to these resources are syntactically different from the user profile tags: $P(u)$ may contain the tag "bloggingstuff" while relevant resources are tagged with "blog", "weblog" or "blogging". To counter such problems we analyze different strategies that further enrich tag-based user profiles and the set of tags T_{source} particularly.

Tag Similarity χ_{sim}**.** This strategy enriches the initial set of tags T_{source} with tags that have high string similarity to one of the tags in T_{source}. In our analysis we apply Jaro-Winkler distance [19] for computing the similarity between two tags. It considers the number of matching characters, number of transformations that would be required to unify the two strings, as well

as the length of both tags and ranges between 0 (no similarity) and 1 (exact match). For example, the Jaro-Winkler distance between "blogging" and "blogging-stuff" is 0.95.

Cross-system Rules χ_{rules}**.** Cross-system user modeling has to deal with heterogeneous vocabularies (cf. Section 3): the overlap of tags between different folksonomies may be low and tags that are popular in one system might be unpopular in another system [2,5]. Cross-system rules enrich tag-based profiles based on association rules deduced from characteristics observed across two systems. These association rules can be phrased as follows.

> *If tag t_a occurs in the personomy $\mathbb{P}_{u,A}$ of user u in folksonomy system A then tag t_b occurs in u's personomy $\mathbb{P}_{u,B}$ in system B.*

In our cross-system user modeling analysis we exploit association rules that are applicable to the tags T_{source} of a given profile, i.e. $t_a \in T_{source}$. We add tags t_b to T_{source} that are generated by those association rules that have the highest confidence while satisfying a minimum support of 0.1.

We use $\chi_{sim}@k$ and $\chi_{rules}@k$ to refer to these strategies that add the top k most appropriate (based on similarity and confidence respectively) tags to T_{source}.

2.3   Weighting Scheme

We compare different approaches to determine the weight $w(u, t)$ associated with a tag $t \in T_{source}$ in the tag-based profile of a specific user u (see Definition 2).

Term frequency TF**.** For the given set of tags T_{source} and a given set of tag assignments Y_{source}, the term frequency corresponds to the relative number of tag assignments in Y_{source} that refer to the tag $t \in T_{source}$.

$$w_{TF}(u,t) = \frac{|\{r \in R|(u,t,r) \in Y_{source}\}|}{|Y_{source}|} \qquad (2)$$

TF and Inverse Document Frequency $TF \times IDF$**.** The inverse document frequency (IDF) can be applied to value the term specificity. From the perspective of the user modeling strategies, IDF refers to the inverse fraction of the number of distinct users that applied a given tag $t \in T_{source}$.

$$w_{TF \times IDF}(u,t) =$$
$$w_{TF}(u,t) \cdot log(\tfrac{|U|}{|\{u_i \in U|(u_i,t,r) \in Y_{source}\}|}) \qquad (3)$$

The weighting schemes thus require a set of tag assignments Y_{source} as input. In accordance to the source of user data, we obtain Y_{source} either from the target personomy $\mathbb{P}_{target}$, from the target folksonomy $\mathbb{F}_{target}$ or from the foreign personomy $\mathbb{P}_{foreign}$. For example, $TF_{\mathbb{F}_{target}}$ considers all tag assignments from the target folksonomy $\mathbb{F}_{target}$.

2.4   Assembling User Modeling Strategies

The actual user modeling strategy for constructing the tag-based profile is built by combining (i) a data source with (ii) some semantic enrichment method, and

Table 1. Example user modeling strategies based on TF weighting scheme, semantic enrichment and different data sources (target folksonomy $\mathbb{F}_{target}$ or foreign personomy $\mathbb{P}_{foreign}$)

Strategy	Source	Enrich	Weighting	Cross?	Description
$F_{\mathrm{T@\mathbb{F}_t,TF@\mathbb{F}_t}}$ *(baseline)*	$T_{\mathbb{F}_{target}}$	–	$TF_{\mathbb{F}_{target}}$	–	Tags and term frequency weights are taken from the target folksonomy. The tag-based profile thus contains the *most popular tags.*
$P_{\mathrm{T@\mathbb{F}_t,TF@\mathbb{P}_f}}$	$T_{\mathbb{F}_{target}}$	–	$TF_{\mathbb{P}_{foreign}}$	+	While the actual tags in the profile are obtained from the target folksonomy, the weights are computed based on the user-specific personomy from a foreign system.
$P_{\mathrm{T@\mathbb{P}_f,TF@\mathbb{F}_t,\chi_s}}$	$T_{\mathbb{P}_{foreign}}$	χ_{sim}	$TF_{\mathbb{F}_{target}}$	+	Tag-based profile contains user-specific tags from foreign personomy enriched with tag similarity. Weights are taken from the global term frequency in $\mathbb{F}_{target}$.
$P_{\mathrm{T@\mathbb{P}_f,TF@\mathbb{P}_f,\chi_r}}$	$T_{\mathbb{P}_{foreign}}$	χ_{rules}	$TF_{\mathbb{P}_{foreign}}$	+	Tag-based profile is constructed based on user-specific personomy as available in the foreign system enriched with cross-system rule.

(iii) a weighting scheme. The semantic enrichment is an optional feature. For example, $P_{\mathrm{T@\mathbb{P}_t,TF@\mathbb{P}_t}}$ corresponds to the tag-based profile proposed in [8,15], which exploits the user-specific personomy in the target system. $P_{\mathrm{T@\mathbb{P}_t,TF@\mathbb{P}_t}}$ would not be applicable in cold-start situations where new users register to a system as the target personomy $\mathbb{P}_{\mathrm{target}}$ would be empty in such situations. Table 1 lists some strategies that can be applied in cold-start situations. For example, $P_{\mathrm{T@\mathbb{P}_f,TF@\mathbb{F}_t,\chi_s}}$ denotes the strategy that utilizes the personomy $\mathbb{P}_{foreign}$ from a foreign system as data source, enriches the profile using tag similarity, and computes weights according to the global term frequency in the target folksonomy.

Based on the different combinations of data sources and weighting schemes and the additional option of semantic enrichment we obtain a powerful framework for generating cross-system user modeling strategies.

3   Analysis of Tag-Based Profiles on the Social Web

Before evaluating the user modeling strategies presented in the previous section, it is important to study the nature of personomies and of the corresponding tag-based profiles ($P_{\mathrm{T@\mathbb{P}_t,TF@\mathbb{P}_t}}$) distributed across the different Social Web systems. In particular, we approach the following research questions.

1. What are the characteristics of the individual tag-based profiles in Twitter, Flickr and Delicious?
2. How do the tag-based profiles of individual users overlap between the different systems?

3.1   Data Collection

To investigate the questions above, we crawled public profiles of 421188 distinct users via the Social Graph API[4], which makes information about connections between user accounts available via its Web service. By exploiting Google profiles

---

[4] http://code.google.com/apis/socialgraph/

Table 2. Tagging statistics of the (a) Twitter-Delicious dataset (1500 users) and (b) Flickr-Delicious dataset (1467 users)

(a) Twitter-Delicious (1500 users)   (b) Flickr-Delicious (1467 users)

	Twitter	Delicious	Aggregated
distinct tags	25668	72901	91515
TAS	80464	619856	700320
distinct tags/user	26.1	180.06	191.88
TAS/user	53.64	413.24	1466.88
tagged resources	57236	124520	181756
resources per user	38.99	91.13	130.12

	Flickr	Delicious	Aggregated
distinct tags	72671	59275	119056
TAS	892378	683665	1576043
distinct tags/user	109.44	189.92	292.63
TAS/user	608.30	466.03	1074.33
tagged resources	166423	109242	272701
resources per user	113.45	85.07	198.50

of users, who interlinked their different online accounts, the API provides the list of accounts associated with a particular user.

For our experiments we were interested in users having accounts at Twitter, Delicious and Flickr. Given the 421188 users, 2007 users linked their Twitter *and* Delicious accounts and 1500 of these users applied tags in both systems. Table 2(a) lists the corresponding tagging statistics of these users. Accordingly, Table 2(b) shows the number of tags, resources, and tag assignments (TAS) of the 1467 users, who linked their Flickr *and* Delicious profiles. We make both datasets available online [1].

3.2   Tag-Based Profiles within Systems

By nature, Twitter is not a typical social tagging system. However, it enables users to annotate posts by means of *hash tags*, which start with the "#" symbol. Table 2(a) indicates that people make use of this tagging feature. On average, each user performed 53.64 (hash) tag assignments on 38.99 Twitter posts using 26.1 distinct (hash) tags. In Delicious, the same users are tagging more frequently with an average of 413.24 tag assignments and an average number of 180.06 distinct tags per user profile. Further, users assign, on average, 4.53 tags to each of their Delicious bookmark while they only attach 1.38 hash tags to their Twitter posts.

The second dataset obtained form those users, who linked their Flickr and Delicious profiles, shows similar statistics for the Delicious profiles (see Table 2(b)). Further, it is interesting to see that users perform, on average, 142.27 tag assignments more in Flickr than they do in Delicious. However, at the same time their Flickr profile contains less distinct tags (109.44) than their Delicious profiles (189.92). The variety of tags in the Delicious profiles is thus higher than the ones in Flickr profiles, which might make them more valuable for personalization (see Section 4).

Figure 1 describes the above observation in detail: less than 20% of the tag-based Flickr profiles contain more than 200 tags and less than 5% of the Twitter profiles contain more than 100 distinct (hash) tags. By contrast, more than 40% of the Delicious profiles have more than 200 tags.
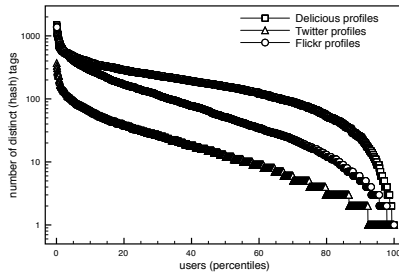
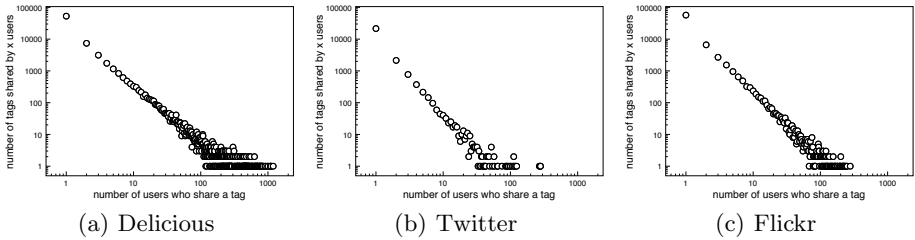**Fig. 1.** Distinct tags per tag-based Twitter, Flickr and Delicious profile

(a) Delicious   (b) Twitter   (c) Flickr

Fig. 2. Homogeneity of tag-based profiles: number of tags that occur in a certain number of (a) Delicious, (b) Twitter, and (c) Flickr profiles

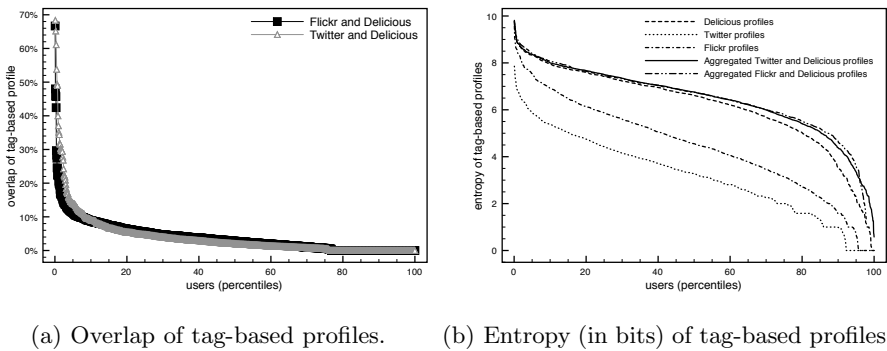(a) Overlap of tag-based profiles.   (b) Entropy (in bits) of tag-based profiles.

Fig. 3. Characteristics of tag-based profiles: (a) overlap of tag-based Twitter and Delicious profiles as well as Flickr and Delicious profiles and (b) entropy (in bits) of tag-based profiles: cross-system user modeling by means of profile aggregation increases entropy of tag-based profiles

Figure 2 describes the tag usage and therefore the homogeneity of the profiles in the three different systems. In all three systems, most tags (e.g., more than 50000 tags in Delicious) only occur in exactly one profile and only a few single tags occur in more than 100 profiles. For example, there exists only one tag, namely "design", that occurs in more than 80% of the Delicious profiles.

3.3   Tag-Based Profiles across Systems

In the context of cross-system user modeling the overlap of tag-based profiles is of particular interest. If the tagging behavior of an individual user differs from system to system then the overlap of the user's tag-based profiles will be small and the corresponding tag-based profiles will probably reveal complementary facets for characterizing the user. Figure 3(a) shows to which degree the tag-based profiles of the individual users overlap between the different systems. We see that the overlap between tag-based Twitter and Delicious profiles of individual users as well as the overlap between the Flickr and Delicious profiles is very small. In fact, for less than 10% of the users, the Twitter or Flickr profiles have an overlap of more than 10% with the tag-based Delicious profiles. This observation indicates that the re-use of tag-based profiles for solving the user modeling challenge posed at the beginning of this paper is not trivial at all and has to be done in an intelligent way. For example, simply re-using a tag-based Twitter profile in Delicious would reflect only a very small part of the user's characteristics in Delicious.

The small overlap of the tag-based profiles implicates that the profiles of an individual user at different Social Web platforms reveal different characteristics of the user. Cross-system user modeling by means of profile aggregation thus allows for more valuable profiles with regards to the information about the user. In Figure 3(b) we compare the entropy of tag-based profiles. Entropy quantifies the information embodied in a tag-based profile $P(u)$, which specifies a weight for each tag $t \in T_{source}$ (see Definition 2), and is computed as follows.

$$entropy(P(u)) = \sum_{t \in T_{source}} p(t) \cdot (-log_2(p(t))) \qquad (4)$$

In Equation 4, $p(t)$ denotes the probability that the tag t was utilized by the corresponding user and can be modeled via the weight $w(u,t)$ from the normalized tag-based profile $\bar{P}(u)$, for which the sum of all weights is equal to 1.

Figure 3(b) shows that tag-based profiles in Delicious have higher entropy than the ones in Twitter and Flickr – even though Flickr features the highest number of tag assignments per user (cf. Table 2(b)). Thus, the variety of tags in Delicious profiles is higher than in Flickr profiles. We further aggregated the profiles of the individual users by accumulating their tag-based profiles from Twitter and Delicious as well as Flickr and Delicious. For both types of profiles entropy increases significantly with respect to the service-specific profiles. Hence, cross-system user modeling based on profile aggregation has significant impact on the entropy of tag-based profiles.

3.4   Synopsis

From our analysis of tag-based profiles $(P_{T@\mathbb{P}_t, TF@\mathbb{P}_t})$ we conclude that users reveal different profile facets in the different systems on the Social Web. Tag-based profiles of same users in different systems overlap only little. For more than 90% of the users, the tag-based profiles obtained from Delicious, Twitter and Flickr overlap to less than 10%. Cross-system user modeling allows for more

valuable profiles. By aggregating tag-based profiles, information gain and entropy improve significantly.

4   Analysis of Cross-System User Modeling Strategies

The above results reveal already benefits of cross-system user modeling for tag-based profiles. We now evaluate the performance of the cross-system user modeling strategies for supporting tag and bookmark recommendations in cold-start situations (see Definition 3) and investigate the following research questions.

1. Which user modeling strategies generate the most valuable tag-based profiles for recommending tags and resources to users in which context?
2. How do the different building blocks of the user modeling strategies (e.g. source of user data) influence the quality of the tag-based profiles?

4.1   Experimental Setup

The actual recommendation algorithm incorporates a user profile $P(u)$, as delivered by the given user modeling strategy, to compute a ranking of items (tags and resources) so that items relevant to the user u appear at the top of the ranking. Therefore, we specify a generic algorithm that can be customized with specific ranking and user modeling strategies.

Generic Recommendation Algorithm *The generic recommendation algorithm recommend(u, s, um) computes a ranked list of entities appropriate to a user u by exploiting a given ranking strategy s and a given user modeling strategy um.*

  1. *$\textbf{Input:}$ ranking strategy s, user modeling strategy um, user u*
  2. *$P(u) = um.modelUser(u)$ (compute user profile)*
  3. *$\tau = s.rank(P(u))$ (rank entities w.r.t. $P(u)$)*
  4. *$\textbf{Output:}$ τ (ranked list of entities)*

As we do not aim to optimize the ranking strategy, but want to investigate the impact of the different cross-system user modeling strategies on the recommendation performance, we deliberately apply lightweight algorithms for ranking the tags and resources respectively.

Tag Recommendation. Regarding the tag recommendation task, our goal is to point the user to tags she has not applied yet, but might be of interest for her and are worthwhile to explore. Hence, we filter the tag-based profile and remove those tags that already occur in the personomy of the user u in the target system ($\mathbb{P}_{u,target}$). The remaining tags are then ranked according to the weight that is associated with the tags in the tag-based profile $P(u)$. If two tags have the same weight then the order of these tags is chosen randomly.

Resource Recommendation. We apply cosine similarity to generate resource recommendations. Given a tag-based user profile $P(u)$, the corresponding

personomy of the user $\mathbb{P}_{u,target} = (T_u, R_u, Y_u)$, the folksonomy of the target system $\mathbb{F}_{target} = (T, R, Y)$, and a set of resources $R_{target} \subseteq R$, which are not explicitly connected to u ($R_u \cap R_{target} = \emptyset$), we first generate tag-based profiles for each resource $r \in R_{target}$. The tag-based resource profile resembles the tag-based user profile specified in Definition 2: $P(r) = \{\{t, w(r,t)\}|t \in T, r \in R_{target}\}$. In accordance to the TF weighting scheme applied by the user modeling strategies, we compute the weights associated with the resources as follows.

$$w_{TF}(r,t) = \frac{|\{u \in U|(u,t,r) \in Y\}|}{|\{u \in U, t_x \in T|(u,t_x,r) \in Y\}|} \tag{5}$$

Given the target folksonomy $\mathbb{F}_{target}$, the weight $w(r,t)$ is thus given by the number of users who assigned tag t to resource r divided by the overall number of tag assignments attached to r. To compute cosine similarity, the tag-based profiles of the user ($P(u)$) and resource ($P(r)$) are represented in a vector space model via $\boldsymbol{u}$ and $\boldsymbol{r}$ respectively where each dimension of these vectors corresponds to a certain tag t that occurs in both profiles. Finally, each resource $r \in R_{target}$ is ranked according its cosine similarity with the user profile representation $\boldsymbol{u}$ so that those resources that have high similarity occur at the top of the ranking.

Evaluation Methodology. We evaluate the quality of the recommender algorithms and therewith the user modeling strategies respectively by means of a *leave-many-out evaluation*. For simulating a cold-start where a new or yet unknown user u is interested in recommendations in system A, we first remove all tag assignments Y_u performed by u in system A from the folksonomy. Each recommender strategy then has to compute a ranked list of recommendations. The quality of the recommendations is measured via *MRR* (Mean Reciprocal Rank), which indicates at which rank the first *relevant* item occurs on average, *S@k* (success at rank k), which stands for the mean probability that a *relevant* item occurs within the top k of the ranking, and *P@k* (precision at rank k), which represents the average proportion of *relevant* items within the top k (cf. [17]). We consider only these items as relevant to which the user u actually referred to in the tag assignments Y_u that were removed before computing the recommendations. For a given user modeling strategy, we ran the experiment for each individual user of our datasets (cf. Section 3). We tested the statistical significance of our results with a two-tailed t-Test where the significance level was set to $\alpha = 0.01$.

4.2   Results: Cold-Start Tag Recommendations

Table 3 gives a detailed overview on the performance of the different user modeling strategies for recommending *tags* in Twitter and Delicious. The best strategies (emphasized) benefit from cross-system user modeling and outperform the other strategies significantly. For example, the precision of the top ten tag recommendations in Delicious based on tags obtained from the user's personomy in Twitter ($T_{\mathbb{P}_{foreign}}$), which are weighted according to the global tag frequencies obtained from the Delicious folksonomy ($TF_{\mathbb{F}_{target}}$), is 94.4% ($P_{T@\mathbb{P}_f, TF@\mathbb{F}_t}$ strategy) and therewith improves by more than 90% over the $P_{T@\mathbb{F}_t, TF@\mathbb{F}_t}$ baseline

Table 3. User modeling performances measured via tag recommendation quality in a cold-start setting in Twitter and Delicious: tags that occur in the tag-based profile are either selected from the folksonomy of the target system ($T_{\mathbb{F}_{target}}$) or from the user-specific personomy in the foreign system ($T_{\mathbb{P}_{foreign}}$) whereas the weighting scheme is term frequency (TF) or TF multiplied by inverse document frequency ($TF \times IDF$), which is either computed based on the target folksonomy ($TF_{\mathbb{F}_{target}}$) or foreign personomy ($TF_{\mathbb{P}_{foreign}}$). "Cross?" denotes whether the corresponding strategy benefits from cross-system user modeling ($+$) or not ($-$). Semantic enrichment (Enrich) is either based on tag similarity (χ_{sim}), cross-system association rules (χ_{rules}) or not applied ($-$).

Strategy	Source	Enrich	Weighting	Cross?	MRR	S@1	S@10	P@5	P@10	P@20
			Twitter (foreign) → Delicious (target):							
$P_{T@\mathbb{P}_f,TF@\mathbb{F}_t}$	$T_{\mathbb{P}_{foreign}}$	−	$TF_{\mathbb{F}_{target}}$	+	**0.979**	96.6	99.6	95.1	94.4	91.4
$P_{T@\mathbb{P}_f,TF@\mathbb{P}_f}$	$T_{\mathbb{P}_{foreign}}$	−	$TF_{\mathbb{P}_{foreign}}$	+	0.631	43.7	97.7	52.4	62.1	72.5
$F_{T@\mathbb{F}_t,TF@\mathbb{F}_t}$	$T_{\mathbb{F}_{target}}$	−	$TF_{\mathbb{F}_{target}}$	−	0.764	69.1	94.5	50.8	49.5	48.4
$P_{T@\mathbb{F}_t,TF@\mathbb{P}_f}$	$T_{\mathbb{F}_{target}}$	−	$TF_{\mathbb{P}_{foreign}}$	+	0.623	51.5	85.3	34.1	28.5	26.4
$P_{T@\mathbb{P}_f,TFxIDF@\mathbb{F}_t}$	$T_{\mathbb{P}_{foreign}}$	−	$TF \times IDF_{\mathbb{F}_{target}}$	+	**0.974**	95.6	99.6	95.3	94.4	91.5
$P_{T@\mathbb{P}_f,TFxIDF@\mathbb{P}_f}$	$T_{\mathbb{P}_{foreign}}$	−	$TF \times IDF_{\mathbb{P}_{foreign}}$	+	0.650	4.5	97.8	52.8	62.1	72.4
$F_{T@\mathbb{F}_t,TFxIDF@\mathbb{F}_t}$	$T_{\mathbb{F}_{target}}$	−	$TF \times IDF_{\mathbb{F}_{target}}$	−	0.244	4.5	88.0	12.5	28.9	33.9
$P_{T@\mathbb{F}_t,TFxIDF@\mathbb{P}_f}$	$T_{\mathbb{F}_{target}}$	−	$TF \times IDF_{\mathbb{P}_{foreign}}$	+	0.608	49.2	85.1	33.4	28.3	26.4
$P_{T@\mathbb{P}_f,TF@\mathbb{F}_t,\chi_s}$	$T_{\mathbb{P}_{foreign}}$	χ_{sim}	$TF_{\mathbb{F}_{target}}$	+	**0.978**	96.4	99.6	94.1	93.8	91.0
$P_{T@\mathbb{P}_f,TF@\mathbb{P}_f,\chi_s}$	$T_{\mathbb{P}_{foreign}}$	χ_{sim}	$TF_{\mathbb{P}_{foreign}}$	+	0.631	43.7	97.7	52.4	62.1	72.6
$P_{T@\mathbb{P}_f,TF@\mathbb{F}_t,\chi_r}$	$T_{\mathbb{P}_{foreign}}$	χ_{rules}	$TF_{\mathbb{F}_{target}}$	+	0.778	69.1	93.3	71.0	68.5	63.1
$P_{T@\mathbb{P}_f,TF@\mathbb{P}_f,\chi_r}$	$T_{\mathbb{P}_{foreign}}$	χ_{rules}	$TF_{\mathbb{P}_{foreign}}$	+	0.624	43.7	96.9	53.2	61.4	71.5
			Delicious (foreign) → Twitter (target):							
$P_{T@\mathbb{P}_f,TF@\mathbb{F}_t}$	$T_{\mathbb{P}_{foreign}}$	−	$TF_{\mathbb{F}_{target}}$	+	0.194	9.7	39.9	10.1	10.7	11.2
$P_{T@\mathbb{P}_f,TF@\mathbb{P}_f}$	$T_{\mathbb{P}_{foreign}}$	−	$TF_{\mathbb{P}_{foreign}}$	+	**0.970**	96.9	96.9	82.1	64.0	42.2
$F_{T@\mathbb{F}_t,TF@\mathbb{F}_t}$	$T_{\mathbb{F}_{target}}$	−	$TF_{\mathbb{F}_{target}}$	−	0.102	5.9	20.4	3.4	3.2	2.9
$P_{T@\mathbb{F}_t,TF@\mathbb{P}_f}$	$T_{\mathbb{F}_{target}}$	−	$TF_{\mathbb{P}_{foreign}}$	+	0.749	74.8	74.8	45.5	29.5	17.0
$P_{T@\mathbb{P}_f,TFxIDF@\mathbb{F}_t}$	$T_{\mathbb{P}_{foreign}}$	−	$TF \times IDF_{\mathbb{F}_{target}}$	+	0.683	60.5	84.4	33.7	27.2	21.6
$P_{T@\mathbb{P}_f,TFxIDF@\mathbb{P}_f}$	$T_{\mathbb{P}_{foreign}}$	−	$TF \times IDF_{\mathbb{P}_{foreign}}$	+	0.386	27.5	60.9	19.9	16.8	14.5
$F_{T@\mathbb{F}_t,TFxIDF@\mathbb{F}_t}$	$T_{\mathbb{F}_{target}}$	−	$TF \times IDF_{\mathbb{F}_{target}}$	−	0.171	8.5	33.7	4.8	5.4	5.8
$P_{T@\mathbb{F}_t,TFxIDF@\mathbb{P}_f}$	$T_{\mathbb{F}_{target}}$	−	$TF \times IDF_{\mathbb{P}_{foreign}}$	+	0.209	11.8	38.0	8.5	6.7	5.0
$P_{T@\mathbb{P}_f,TF@\mathbb{F}_t,\chi_s}$	$T_{\mathbb{P}_{foreign}}$	χ_{sim}	$TF_{\mathbb{F}_{target}}$	+	0.194	9.7	39.7	10.1	10.7	11.2
$P_{T@\mathbb{P}_f,TF@\mathbb{P}_f,\chi_s}$	$T_{\mathbb{P}_{foreign}}$	χ_{sim}	$TF_{\mathbb{P}_{foreign}}$	+	**0.970**	96.9	96.9	82.1	64.0	42.1
$P_{T@\mathbb{P}_f,TF@\mathbb{F}_t,\chi_r}$	$T_{\mathbb{P}_{foreign}}$	χ_{rules}	$TF_{\mathbb{F}_{target}}$	+	0.096	5.9	18.8	3.9	5.2	4.2
$P_{T@\mathbb{P}_f,TF@\mathbb{P}_f,\chi_r}$	$T_{\mathbb{P}_{foreign}}$	χ_{rules}	$TF_{\mathbb{P}_{foreign}}$	+	**0.969**	96.9	96.9	83.8	68.3	47.2

strategy, which does not make use of cross-system user modeling and achieves a precision of 49.5.

This improvement is even significantly higher for recommending tags in Twitter where the best cross-system user modeling strategy ($P_{T@\mathbb{P}_f,TF@\mathbb{P}_f,\chi_r}$) achieves 68.3 for $P@10$ in comparison to 5.4, achieved by the best non-cross-system strategy. Overall, tag recommendations in Twitter seem to be more difficult: while the best strategies allow for high success values (e.g., $S@1 > 95$), it requires advanced user modeling methods like the application of cross-system association rules (χ_{rules}) to obtain high precisions for $P@10$ and $P@20$. Possible explanations for this can be derived from Section 3: Twitter user profiles are more sparse than the corresponding Delicious profiles and exhibit a lower entropy. This might also explain why $TF_{\mathbb{P}_{foreign}}$, i.e. weighting based on tag-based profile in Delicious, is more successful in Twitter than $TF_{\mathbb{F}_{target}}$, the Twitter-based weighting scheme. Hence, when applying the user modeling strategies in a cross-system setting, it is important to consider the general characteristics of the tagging data available in the different systems (as we did in Section 3). For example, the consideration

(a) Tag recommendations. (b) Resource recommendations.
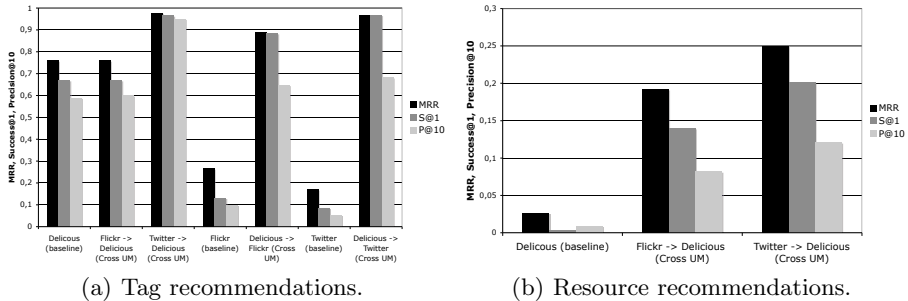
Fig. 4. Comparison of different cross-system user modeling settings in the context of (a) tag recommendations and (b) resource recommendations. The baseline strategies follow the $P_{\text{T@}\mathbb{F}_t,\text{TF@}\mathbb{F}_t}$ approach, $P_{\text{T@}\mathbb{P}_f,\text{TF@}\mathbb{P}_f}$ is applied for settings when the target is Delicious and $P_{\text{T@}\mathbb{P}_f,\text{TF@}\mathbb{P}_f,\chi_r}$ is used for recommendations in Flickr and Twitter.

of Delicious profile data leads to higher improvements than the consideration of Twitter profile data. However, independent from the setting we see that cross-system user modeling has significant impact on the recommendation quality.

Our evaluations on the Flickr-Delicious dataset confirmed the results listed in Table 3 with two remarkable differences that are summarized in Figure 4(a): (1) Flickr-based cross-system user modeling (*Flickr → Delicious (Cross UM)*) for tag recommendations in Delicious improves the quality only slightly (no significant difference) and (2) in the context of tag recommendations in Flickr, cross-system user modeling leads again to tremendous improvements, however, the quality gain is not as high as for the *Delicious → Twitter* setting. The absolute tag recommendation performances are however high for all cross-system settings and work best for the interplay of Delicious and Twitter (e.g. $MRR > 0.95$) as depicted in Figure 4(a). Overall, we conclude that the significant improvements of the tag recommendation performance further support the utility of cross-system user modeling.

4.3   Results: Cold-Start Resource Recommendations

Table 4 overviews the performance of the different user modeling strategies for cold-start bookmark recommendations in Delicious. This task is more difficult than the tag recommendation task as the items, which are recommended to the user, are not directly connected to the user, but indirectly via the tags [16]. Further, the fraction of relevant items is much smaller: given an average number of 91.3 relevant bookmarks per user, the probability to randomly select a resource, which will be bookmarked by the user, from the set of 124520 available resources is 0.0007 ($P@1$). With this in mind, it is interesting to see that the cross-system user modeling strategies, which fully exploit the Twitter personomies (source selection and weighting is based on $\mathbb{P}_{foreign}$), succeed for recommending Delicious bookmarks in a cold-start setting and perform significantly better than all other strategies (see Table 4).

Table 4. User modeling performances measured via Delicious bookmark recommendation quality in a cold-start setting with Twitter: tags that occur in the tag-based profile are either selected from the folksonomy of Delicious ($T_{\mathbb{F}_{target}}$) or from the user-specific personomy in Twitter ($T_{\mathbb{P}_{foreign}}$) whereas the weighting scheme is the term frequency (TF) or TF multiplied by inverse document frequency ($TF \times IDF$), which is either computed based on the target Delicious folksonomy ($TF_{\mathbb{F}_{target}}$) or foreign personomy available in Twitter ($TF_{\mathbb{P}_{foreign}}$). "Cross?" denotes whether the corresponding strategy benefits from cross-system user modeling ($+$) or not ($-$). Semantic enrichment (Enrich) is either based on tag similarity (χ_{sim}), cross-system association rules (χ_{rules}) or not applied ($-$).

Strategy	Source	Enrich	Weighting	Cross?	MRR	S@1	S@10	P@5	P@10	P@20
$P_{T@\mathbb{P}_f,TF@\mathbb{F}_t}$	$T_{\mathbb{P}_{foreign}}$	–	$TF_{\mathbb{F}_{target}}$	+	0.026	0.3	5.6	0.7	0.6	0.5
$P_{T@\mathbb{P}_f,TF@\mathbb{P}_f}$	$T_{\mathbb{P}_{foreign}}$	–	$TF_{\mathbb{P}_{foreign}}$	+	**0.191**	**14.4**	**26.0**	**11.7**	**9.3**	**7.7**
$P_{T@\mathbb{F}_t,TF@\mathbb{F}_t}$	$T_{\mathbb{F}_{target}}$	–	$TF_{\mathbb{F}_{target}}$	–	0.026	0.3	7.8	1.0	0.9	0.6
$P_{T@\mathbb{F}_t,TF@\mathbb{P}_f}$	$T_{\mathbb{F}_{target}}$	–	$TF_{\mathbb{P}_{foreign}}$	+	0.055	2.3	11.1	2.1	1.9	2.1
$P_{T@\mathbb{P}_f,TFxIDF@\mathbb{F}_t}$	$T_{\mathbb{P}_{foreign}}$	–	$TF \times IDF_{\mathbb{F}_{target}}$	+	2.7	0.5	3.9	0.6	0.4	0.6
$P_{T@\mathbb{P}_f,TFxIDF@\mathbb{P}_f}$	$T_{\mathbb{P}_{foreign}}$	–	$TF \times IDF_{\mathbb{P}_{foreign}}$	+	**0.244**	**18.9**	**32.2**	**14.5**	**11.6**	**9.4**
$P_{T@\mathbb{F}_t,TFxIDF@\mathbb{F}_t}$	$T_{\mathbb{F}_{target}}$	–	$TF \times IDF_{\mathbb{F}_{target}}$	–	0.006	0.1	1.5	0.1	0.1	0.1
$P_{T@\mathbb{F}_t,TFxIDF@\mathbb{P}_f}$	$T_{\mathbb{F}_{target}}$	–	$TF \times IDF_{\mathbb{P}_{foreign}}$	+	0.053	2.3	10.9	2.0	2.3	2.2
$P_{T@\mathbb{P}_f,TFxIDF@\mathbb{F}_t,\chi s}$	$T_{\mathbb{P}_{foreign}}$	χ_{sim}	$TF \times IDF_{\mathbb{F}_{target}}$	+	0.026	0.5	3.9	0.5	0.4	0.6
$P_{T@\mathbb{P}_f,TFxIDF@\mathbb{P}_f,\chi s}$	$T_{\mathbb{P}_{foreign}}$	χ_{sim}	$TF \times IDF_{\mathbb{P}_{foreign}}$	+	**0.244**	**18.7**	**32.4**	**14.5**	**11.7**	**9.4**
$P_{T@\mathbb{P}_f,TFxIDF@\mathbb{F}_t,\chi r}$	$T_{\mathbb{P}_{foreign}}$	χ_{rules}	$TF \times IDF_{\mathbb{F}_{target}}$	+	0.01	0.1	1.3	0.1	0.2	0.2
$P_{T@\mathbb{P}_f,TFxIDF@\mathbb{P}_f,\chi r}$	$T_{\mathbb{P}_{foreign}}$	χ_{rules}	$TF \times IDF_{\mathbb{P}_{foreign}}$	+	**0.252**	**20.1**	**33.2**	**14.9**	**12.1**	**10.0**

While the TF weighting scheme is best applicable in the tag recommendation context, computing weights based on $TF \times IDF$ and $TF \times IDF_{\mathbb{P}_{foreign}}$ in particular perform best for the Delicious bookmark recommendations. We further discover that the semantic enrichment by means of cross-system association rules, that indicate which tags are likely to occur in the tag-based Delicious profile if a user applied certain tags in Twitter, has significant impact on the quality of the generated profiles. The difference in performance of $P_{T@\mathbb{P}_f,TFxIDF@\mathbb{P}_f,\chi r}$, which is again the best strategy, with respect to the other strategies is statistically significant.

These findings can also be confirmed on the setting of Flickr and Delicious. Figure 4(b) compares the performance of the cross-system user modeling strategy $P_{T@\mathbb{P}_f,TFxIDF@\mathbb{P}_f,\chi r}$ with the best baseline strategy that does not consider external profile information. Exploiting Twitter for resource recommendations in Delicious is more appropriate than the consideration of Flickr profiles. For example, given the Twitter-based profile, the probability to recommend a relevant resource ($S@1$) is 0.2 in comparison with 0.14 for the Flickr-based profiles. Further, with Twitter-based cross-system user modeling we achieve a precision ($P@10$) of 0.12, which is more than 10 times higher than the baseline strategy.

4.4   Synopsis

For both tag and resource recommendations, we conclude that cross-system user modeling strategies generate significantly more valuable tag-based profiles than those strategies which do not consider external profile information. We see that the interplay between Twitter and Delicious is more successful than the one of

Flickr and Delicious. The selection of the building blocks for composing an appropriate cross-system user modeling strategy impacts the quality significantly. For example, in the resource recommendation context those user modeling strategies perform best that fully exploit personomies from other systems (source selection and weighting is based on $\mathbb{P}_{foreign}$).

5   Conclusions

In this paper we analyzed strategies for modeling users across Social Web system boundaries to infer insights for engineers of Web applications that aim for personalization in cold-start situations. We investigated the characteristics of tag-based profiles that result from tagging activities in Flickr, Twitter and Delicious, and discovered that the tag-based profiles which a same user has at different platforms only overlap little. Cross-system user modeling by means of aggregating the user-specific profiles from the different platforms results in significantly more valuable profiles with respect to entropy. Within the scope of tag and resource recommendations in cold-start settings we further revealed that cross-system user modeling strategies have significant impact on the performance of the recommendation quality. For example, user modeling based on external personomies results in a more than 10 times higher precision for recommending Delicious bookmarks.

References

1. Abel, F., Araújo, S., Gao, Q., Houben, G.J., Tao, K.: Supporting website: datasets and further details (2011), `http://wis.ewi.tudelft.nl/icwe2011/um/`
2. Abel, F., Henze, N., Herder, E., Krause, D.: Interweaving public user profiles on the web. In: De Bra, P., Kobsa, A., Chin, D. (eds.) UMAP 2010. LNCS, vol. 6075, pp. 16–27. Springer, Heidelberg (2010)
3. Aroyo, L., Dolog, P., Houben, G.J., Kravcik, M., Naeve, A., Nilsson, M., Wild, F.: Interoperability in pesonalized adaptive learning. J. Educational Technology & Society 9 (2), 4–18 (2006)
4. Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., Su, Z.: Optimizing Web Search using Social Annotations. In: Proc. of 16th Int. World Wide Web Conference (WWW 2007), pp. 501–510. ACM Press, New York (2007)
5. Bischoff, K., Firan, C., Paiu, R., Nejdl, W.: Can All Tags Be Used for Search? In: Proc. of Conf. on Information and Knowledge Management (CIKM 2008). ACM, New York (2008)
6. Carmagnola, F., Cena, F.: User identification for cross-system personalisation. Information Sciences: an International Journal 179(1-2), 16–32 (2009)
7. Dellschaft, K., Staab, S.: An epistemic dynamic model for tagging systems. In: Brusilovsky, P., Davis, H.C. (eds.) Proceedings of the 19th ACM Conference on Hypertext and Hypermedia (HT 2008), pp. 71–80. ACM, Pittsburgh (2008)
8. Firan, C.S., Nejdl, W., Paiu, R.: The Benefit of Using Tag-based Profiles. In: Proc. of 2007 Latin American Web Conference (LA-WEB 2007), pp. 32–41. IEEE Computer Society, Washington, DC, USA (2007)

9. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. Journal of Information Science 32(2), 198–208 (2006)
10. Heckmann, D., Schwartz, T., Brandherm, B., Schmitz, M., von Wilamowitz-Moellendorff, M.: Gumo – The General User Model Ontology. In: Ardissono, L., Brna, P., Mitrović, A. (eds.) UM 2005. LNCS (LNAI), vol. 3538, pp. 428–432. Springer, Heidelberg (2005)
11. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: Search and ranking. In: Proc. of the 3rd European Semantic Web Conference (ESWC 2010), Budva, Montenegro, pp. 411–426. Springer, Heidelberg (2006)
12. Jameson, A.: Adaptive interfaces and agents. In: The HCI Handbook: Fundamentals, Evolving Technologies and Emerging Applications, pp. 305–330 (2003)
13. Kobsa, A.: Generic user modeling systems. User Modeling and User-Adapted Interaction 11(1-2), 49–63 (2001)
14. Mehta, B.: Cross System Personalization: Enabling personalization across multiple systems. VDM Verlag, Saarbrücken (2009)
15. Michlmayr, E., Cayzer, S.: Learning User Profiles from Tagging Data and Leveraging them for Personal(ized) Information Access. In: Proc. of the WWW 2007 Workshop on Tagging and Metadata for Social Information Organization (2007)
16. Sen, S., Vig, J., Riedl, J.: Tagommenders: connecting users to items through tags. In: Quemada, J., León, G., Maarek, Y.S., Nejdl, W. (eds.) Proceedings of the 18th International Conference on World Wide Web (WWW 2009), pp. 671–680. ACM, New York (2009)
17. Sigurbjörnsson, B., van Zwol, R.: Flickr tag recommendation based on collective knowledge. In: Proc. of 17th Int. World Wide Web Conference (WWW 2008), pp. 327–336. ACM Press, New York (2008)
18. Szomszor, M., Alani, H., Cantador, I., O'Hara, K., Shadbolt, N.: Semantic modelling of user interests based on cross-folksonomy analysis. In: Sheth, A.P., Staab, S., Dean, M., Paolucci, M., Maynard, D., Finin, T.W., Thirunarayan, K. (eds.) ISWC 2008. LNCS, vol. 5318, pp. 632–648. Springer, Heidelberg (2008)
19. Winkler, W.E.: The State of Record Linkage and Current Research Problems. Technical report, Statistical Research Division, U.S. Census Bureau (1999)