I Tag, You Tag: Translating Tags for Advanced User Models

Robert Wetzker
Technische Universität Berlin
Berlin, Germany
robert.wetzker@dai-labor.de

Carsten Zimmermann
University of San Diego
San Diego, USA
zimmermann@sandiego.edu

Christian Bauckhage
Fraunhofer IAIS
Bonn, Germany
christian.bauckhage@iais.fraunhofer.de

Sahin Albayrak
Technische Universität Berlin
Berlin, Germany
sahin.albayrak@dai-labor.de

ABSTRACT

Collaborative tagging services (folksonomies) have been among the stars of the Web 2.0 era. They allow their users to label diverse resources with freely chosen keywords (tags). Our studies of two real-world folksonomies unveil that individual users develop highly personalized vocabularies of tags. While these meet individual needs and preferences, the considerable differences between personal tag vocabularies (personomies) impede services such as social search or customized tag recommendation. In this paper, we introduce a novel user-centric tag model that allows us to derive mappings between personal tag vocabularies and the corresponding folksonomies. Using these mappings, we can infer the meaning of user-assigned tags and can predict choices of tags a user may want to assign to new items. Furthermore, our translational approach helps in reducing common problems related to tag ambiguity, synonymous tags, or multilingualism. We evaluate the applicability of our method in tag recommendation and tag-based social search. Extensive experiments show that our translational model improves the prediction accuracy in both scenarios.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval models; H.3.5 [**Online Information Services**]: Web-based services

General Terms

Algorithms, Human Factors

Keywords

user modeling, tagging, folksonomies, tag recommendation, social search

WSDM'10, February 4–6, 2010, New York City, New York, USA.

1. INTRODUCTION

On-line services, such as Delicious[1], Flickr[2], or LibraryThing[3], have evolved into important and popular content collection, management, and sharing tools. They allow their users to label content by assigning freely chosen keywords (tags). In such open environments, tagging has proven a powerful alternative to existing top-down categorization techniques, such as taxonomies or predefined dictionaries, that lack flexibility and are generally expensive in their creation and maintenance. In contrast, tagging allows users to choose the labels that match their real needs, tastes, or language, which reduces the required cognitive efforts. Community services that center around the tagging of resources are called *folksonomies* [27]. These communities have become a valuable source of information, since they bundle the interests, preferences, or contributions of thousand or millions of users.

Folksonomies come in two forms, depending on the underlying tagging rights [26]. *Narrow* folksonomies restrict the bookmarking and thus tagging of resources to only a limited number of users. A common example of this type of folksonomy is Flickr, where photos are labeled by their owner. In this paper, we will instead focus on *broad* folksonomies, such as Delicious or Bibsonomy[4], where items can be bookmarked and tagged by the entire community, such that the tagging activity of one user can be related to the activity of other users. The difference in tagging rights is also expected to have a severe impact on the observed motivations behind tagging. Whereas users of narrow folksonomies may try to promote their content by assigning common tags, i.e. tags other users will most likely search for, this motivation is less present in broad folksonomies, where users tag other users' content. As a result, users of broad folksonomies have a limited interest in the social aspects of tagging, but will assign tags for personal content retrieval. In consequence, we observe these users developing their own tag vocabularies, more commonly known as *personomies* [11], which may differ drastically from those of other users.

A user who wants to structure large collections by tags

[1]http://delicious.com
[2]http://www.flickr.com
[3]http://www.librarything.org
[4]http://www.bibsonomy.org

is not as free as it seems. Instead, there exists an underlying interest in a stable and proven set of category tags for more efficient content retrieval at a later stage. Accordingly, we observe that individual tagging habits remain relatively stable if long-term effects, i.e. changing user interests, can be ignored. Apparently, this trend is further strengthened by the tag recommendation functionalities that all popular tagging sites have introduced to their services. However, an individual user's interest in a converging set of category tags is not a social phenomenon. Instead, we observe that even if two users bookmark conceptually similar content, the tags they prefer for categorizing this content are likely to differ. It simply appears to be a matter of personal taste which category tags a user chooses among existing synonymous forms, singular or plural forms, abbreviations, etc. Also, we find that individual tag vocabularies often reflect the multilingualism of users or their different levels of expertise.

From an information retrieval perspective, the fact that tags are highly personalized may be ignored on a macroscopic scale, where cumulative effects dominate over local heterogeneities. In fact, our studies reveal that resources in broad folksonomies develop a characteristic and stable tag spectrum independent of user-level differences. Consequently, tags have been successfully used to calculate the similarity between items [14, 16, 21] as well as for item clustering and concept extraction [2, 8, 14]. However, effects due to customized tag assignments (TAS) cannot be neglected for applications and scenarios such as personalized tag-based search, tag recommendation, or user interest discovery, which focus on the needs of individual users.

1.1 Our contribution

We study the user-level tagging behavior in two representative broad folksonomies and provide evidence that users follow different tagging strategies. We present a novel, user-centric tag model (UCTM) that overcomes the problems this diversity induces in many potential application areas, including tag recommendation, user interest discovery, and tag-based search. Our model infers the semantics of tags based on who applied them and by looking at the items they have been assigned to. By studying these contexts and their co-occurrence with the tags chosen by other users, we then derive translational mappings between individual personomy tags and the corresponding folksonomy. These mappings help us to overcome common problems related to tag ambiguity, synonymous tags, and biases towards singular or plural forms. The translation of personomy tags even allows us to account for multilingual communities. Furthermore, by translating user tags to the language of the community, we can deduce the meaning of a tag based on who assigned it and design more robust user models that overcome inter-user discrepancies in tagging. On the other hand, translating from the folksonomy language to a user's individual language enables us to predict the concepts (tags) that a user will associate with an item based on the associations of other users. We evaluate the applicability of our translational approach in two user-centric scenarios, that we describe in the next section.

1.2 Application Scenarios

Tag recommendation.

Tag recommenders support a user during the posting process by suggesting potentially relevant tags. In addition to their positive effect on usability, they are effective tools to limit global and local tag divergence by lowering the ratio of misspellings and increasing the likelihood of tag reassignments. We consider the scenario of graph-based tag recommendation, where no additional information about the underlying items, tags, or users is given. Our approach allows us to translate community tags previously assigned to a resource into a user's own tagging language, producing more accurate predictions. Tag recommendation is probably the best-researched recommendation scenario in the context of folksonomies, see e.g. [7, 10, 12, 22, 24, 25]. This scenario therefore allows us the evaluation our approach with respect to previously presented methods.

Social search.

Folksonomies have become important sources for content search, since they aggregate the interests of thousands or millions of users and index content by tags. Our second application scenario considers a user who searches for interesting items related to a tag from her personomy. In this use case, we assume that the user wants to find relevant content related to topics she has been interested in before. Our goal is to simplify this task. Instead of having to guess how other users labeled relevant content, our solution allows the user to rely solely on her own tag vocabulary for search. The challenges in this scenario are not only the personalized retrieval of interesting items, but also the identification of the individual search intention behind a given tag. We will show that this intention can be inferred by our translation approach, yielding more useful search results. Although we limit our discussion here to a search scenario, we believe that our findings also apply to related tasks such as the topic-aware recommendation of items or the topic-aware detection and indication of trends.

The remainder of this paper is structured as follows. We begin with a discussion of related work and provide a description of the folksonomy datasets we use for our studies and evaluations. We then analyze these datasets in order to verify the assumptions that motivate our approach. This is followed by the description of our translational model and its evaluation in the application scenarios described above.

2. RELATED WORK

Recent years have seen an increasing number of publications on different aspects of folksonomies. Among the first publications are [6], [17], and [18] that provide an initial overview on the structure and dynamics of collaborative tagging systems and start modeling folksonomies as tripartite hypergraphs [18].

2.1 Tag models

The authors of [6] investigate the user intentions behind tagging. Of the tags they analyzed a vast majority referred to the topic, type, or owner of the underlying content. These results have been confirmed and extended to various types of folksonomies by Bischoff et al. [3]. Both studies, [6] and [3], do not investigate the differences in user-level tagging strategies. The authors of [19] propose an approach to tag translation for cross-language retrieval. Tags are considered valid translations if they expose similar global tag co-occurrence patterns. The contextualization of tags for improved tag disambiguation was proposed by Yeung et al. [2]. The authors present a co-occurrence measure that considers two resources as similar if a user has labeled them with the same tag. They report that this user-centric measure is better adapted for tag disambiguation than other co-occurrence measures. Further research on various tag co-occurrence distributions is presented in [5, 16, 20]. Li et al. [14] note that users of the Delicious bookmarking service use personalized vocabularies for the same URLs. They further argue that these differences can be neglected when looking only at URL tags due to cumulative effects. Both observations inspire our translation approach, which uses the tag co-occurrences between personomies and the folksonomy within the tagged resources in order to create mappings. However, we believe that the user-level differences in tagging can only be neglected in non user-centric scenarios.

2.2 Tag recommendation

One of the first works on graph-based tag recommendation in folksonomies is [12]. The authors compare the performance of co-occurrence-based tag recommenders and more advanced methods like the FolkRank algorithm [11]. They report only minor improvements of the FolkRank method over the co-occurrence approaches. Instead of reducing the folksonomy hypergraph to bipartite co-occurrence graphs, Symeonidis et al. [25] propose to decompose the full folksonomy tensor and estimate missing values using Higher Order SVD (HOSVD). The authors claim reasonably good results on the task of tag recommendation. Rendle et al. [22] report an even better performance when using ranking-oriented tensor factorization methods. Most similar to our work is an approach by Lipczak [15]. The proposed tag recommender weights the tags of each user's personomy with respect to community tags using a lexicon-based approach, but no further details about the mappings are provided. Other, less related work on tag recommendation is [7, 10, 24].

2.3 Social search

Because of the user-driven aggregation and labeling of content, folksonomies are valuable sources for content retrieval. Heymann et al. [9] investigate the potential of social bookmarking services, such as Delicious, for enhanced web search. They show that there exists a reasonably high overlap between search query terms and tags found in Delicious. Similar studies are presented by Krause et al. [13]. Zanardi et al. [32] approach the problem of content recommendation in folksonomies by a collaborative filtering approach where the similarity between users is determined by tag overlap. Furthermore, the authors expand tag queries by including closely related tags. Wu et al. [30] develop a probabilistic latent topic model for folksonomies. Their model returns items that fall into the same latent topic(s) as the user query. Similar to this approach, Wetzker at al. [28] propose a hybrid item recommender that merges collaborative filtering and tag-based models into a unified representation. Their results show an improved performance of standard collaborative filtering techniques when adding tags. Another hybrid approach to item recommendation is presented by Xu et al. [31], who design a personalized search framework that ranks documents by their similarity to a user and the input query in the tag space. Finally, the universality of the FolkRank algorithm also allows its deployment in a social search scenario, as demonstrated by Hotho et al. [11]. None of these existing works consider the meaning of a tag to be user specific. Instead, it is generally assumed that the meaning of a tag is intrinsic.

3. DATASETS

We conduct our analysis and evaluate our model based on large snapshots of two popular social bookmarking services.

Delicious is probably the best researched folksonomy to date. Its users can centrally save and organize their bookmark collections on the Web. Our Delicious dataset is a subset of the one presented in [29], which is publicly available upon request. To avoid spam-related artifacts, we remove users with a very high concentration of URLs that point to the same domain(s). This filters highly influential spammers who intend to promote only a limited number of domains by posting many URLs. Of the spam-filtered corpus, we only keep the bookmarks recorded in the first 16 months from September 2003 until December 2004.

Bibsonomy is another highly researched folksonomy where users collect and annotate URLs as well as publications. We use the Bibsonomy snapshot that was made available during the 2009 ECML PKDD tag recommendation challenge. We simplify the scenario by fusing URLs and publications into a single item set, as previously done by other authors [11, 12, 22, 25]. The dataset does not require any spam filtering, since spam was manually removed by the Bibsonomy team.

From both datasets we only consider the actual network between users, items, and tags and ignore all additional data, such as content descriptions or bibliographic information in the case of publications. Our user-level analysis of tagging is based on the full folksonomy graphs. For the tag recommendation and the search scenario we instead follow common practice and perform our evaluations on p-core versions of the original graphs, with p set to 5. The p-core of a folksonomy graph is its largest subgraph where all users, items, and tags appear in at least p bookmarks of the subgraph [12]. The resulting global dataset statistics are shown in Table 1.

Table 1: Statistics for both datasets and p-cores.

	p	TAS	Posts	Items	Tags	Users
Bibsonomy	1	1,401,104	421,928	378,378	93,756	3,617
	5	55,008	13,973	1,755	1,584	358
Delicious	1	3,906,207	1,909,687	982,356	140,645	25,966
	5	1,404,101	656,159	42,058	13,204	13,757

4. THE NEED FOR TAG TRANSLATION

Our translational approach is motivated by several assumptions. We examine the validity of these assumptions by analyzing the available data.

A1: People tag for personal later retrieval.

The intention to save content for possible later use is a general characteristic of folksonomies [27]. To reach an optimal retrieval efficiency, folksonomy members will tend to avoid the usage of synonymous tags. Furthermore, one might expect that folksonomy users are less interested in the social effects of their tagging behavior, resulting in individual tagging strategies. To exemplify this behavior, Figure

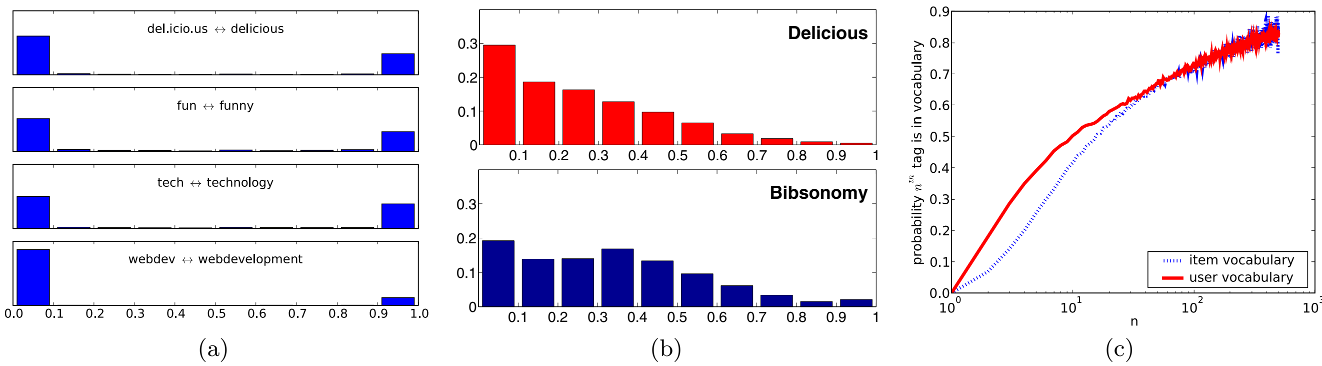


Figure 1: (a) Intra-personomy distribution of synonymous tags in Delicious. The x-axis measures the ratio at which a user chooses the right tag from both choices. The histograms emphasize that users avoid mixing equivalent tags, since this would complicate later resource retrieval. (b) Cosine similarities between the tags that two randomly selected users assigned to the same resources. Only user pairs that shared at least 5 items have been included in the evaluation. The plot shows that the tags which two users apply to the same items do not have to overlap. Hence, users who obviously share the same interests by bookmarking the same content do not have to appear to be similar in the tag space. (c) Probability that a new tag exists in a user's personomy or in the item vocabulary as a function of the number of previous tag assignments (TAS) by each user and to each resource (Delicious). The tag coverage constantly increases with growing vocabularies, indicating converging tag distributions.

1(a) investigates to what degree Delicious users mix synonymous tags, such as "webdev" versus "webdevelopment". The x-axis shows the ratio at which a user selects one of the two synonyms. The figure illustrates that users tend to avoid using both tags in parallel, but rather stick to their initially chosen alternative. Similar observations can be made for Bibsonomy. Table 2 lists the tags which two Bibsonomy users assigned to the same resources as well as the Top 5 community tags. Both users have developed standardized sets of category tags, avoiding the mixture of equivalent tags, such as "collaborative" versus "collaboration" or "folksonomy" versus "folksonomies". These categories do not necessarily have to overlap with the community opinion, but may satisfy more personal classification needs, as demonstrated by *User237*.

Further support for our assumption comes from the observation that 15 percent of all Delicious users in our dataset group tags to so called *tag bundles*. This feature of Delicious lets users define categories and list the personomy tags that should fall into each category, providing users with an additional top-down classification technique. Moreover, since tag bundles behave like meta-tags they can help find previously bookmarked items by category. The rather frequent use of this technique further highlights the retrieval motivation behind tagging.

A2: People tag differently.

Table 2 also highlights inter-user differences in tagging strategies. While *User107* has a tendency towards commonly assigned tags, we find the second user (*User237*) to have developed a personalized set of labels including tags, such as "diplomathesis" or "closelyrelated", that rather describe the user's relation to the respective content than the content itself. Figure 1(b) further emphasizes the user-centric nature of tagging by looking at the cosine similarity between the tags two randomly selected users assigned to the same resources. The evaluation only includes pairs of users who share at least five items. Independent of the dataset, about 20 to 30 percent of all pairs of users labeled the same resources in such a different manner that there exists nearly no overlap between the assigned tags. The divergence of categorization strategies likely even exceeds the one shown in Figure 1(b), because, as we will discuss in the following, many overlapping tags describe a specific item and are not part of any categorization scheme. Therefore, even in cases where two users obviously share common interests by bookmarking the same content, they do not necessarily appear similar in the tag space. Deriving user similarities by simple personomy comparison, cf. e.g. [23, 32], will hence lead to suboptimal results. We therefore propose first to translate the personomies of users into a unified global representation before taking further steps.

A3: People share a common understanding about content.

A prerequisite of our model is the relative stability of tag distributions from an item perspective. Figures 2(a)-(c) look at how the tag distributions of three exemplary URLs develop over time. All plots show a convergence after only a few bookmarking events. The collaborative labeling of content thus leads to the emergence of a characteristic and stable tag spectrum despite of user-level discrepancies. This spectrum is dominated by a minority of very frequent tags, with the most popular tag being assigned in more than 25 percent of all cases independent of the URL. These dominant tags are not item specific, describing the name of the URL or service, but rather describe the primal concern of the resource, such as "news" (a), "reference" (b), or "css" (c). On the other hand, looking at the node degrees unveils a majority of tags that appear only once. This inequality between dominant and infrequent tags seems even more pronounced than in power law distributions, that have been comprehensively reported for various network types including folksonomies [11, 14, 20]. We attribute this bias to the

Table 2: Tags assigned to publications by the community (Top 5) and two exemplary users (Bibsonomy).

Publication title	**Community**	**User107**	**User237**
Ontologies Are Us: A Unified Model of Social Networks and Semantics	folksonomy, ontology, ontologies, social, socialnetworks	social, ontology, network, sna	socialnetworks, ontologies, diplomathesis, closelyrelated
Harvesting social knowledge from folksonomies	folksonomy, tagging, social, knowledge, folksonomies	social, tagging, folksonomy, knowledge, collaborative	diplomathesis, closelyrelated
Towards the Semantic Web: Collaborative Tag Suggestions	tagging, folksonomy, recommender, collaborative, web20	tagging, collaborative	tagging, folksonomy, diplomathesis, closelyrelated
The PageRank Citation Ranking: Bringing Order to the Web	pagerank, search, ranking, ir, web	search, ir, ranking, pagerank	diplomathesis, eventuallyuseful
Automated Tag Clustering: Improving search and exploration in the tag space	clustering, tagging, search, folksonomy, toread	search, folksonomy, clustering, tags, collaborative, tag, filtering, exploration	diplomathesis, closelyrelated
Integrating Folksonomies with the Semantic Web	folksonomy, semanticweb, tagging, semantic, clustering	semanticweb, tagging, folksonomy, projet	toread
The Dynamics and Semantics of Collaborative Tagging	tagging, semantics, ontology, folksonomy, dynamics	tagging, collaborative, filtering	ontology, diplomathesis, closelyrelated, semantics, collaborativetagging
Exploring social annotations for the semantic web	semanticweb, tagging, folksonomy, social, annotation	semanticweb, social, annotation	semanticweb, tagging, diplomathesis, folksonomy, closelyrelated
Collaborative tagging as a tripartite network	tagging, folksonomy, network, collaboration, clustering	tagging, network, collaborative, projet, projtags	diplomathesis, folksonomybackground
Folksonomy as a Complex Network	folksonomy, tagging, network, folksonomybackground, social	folksonomy, network, complex	diplomathesis, folksonomybackground

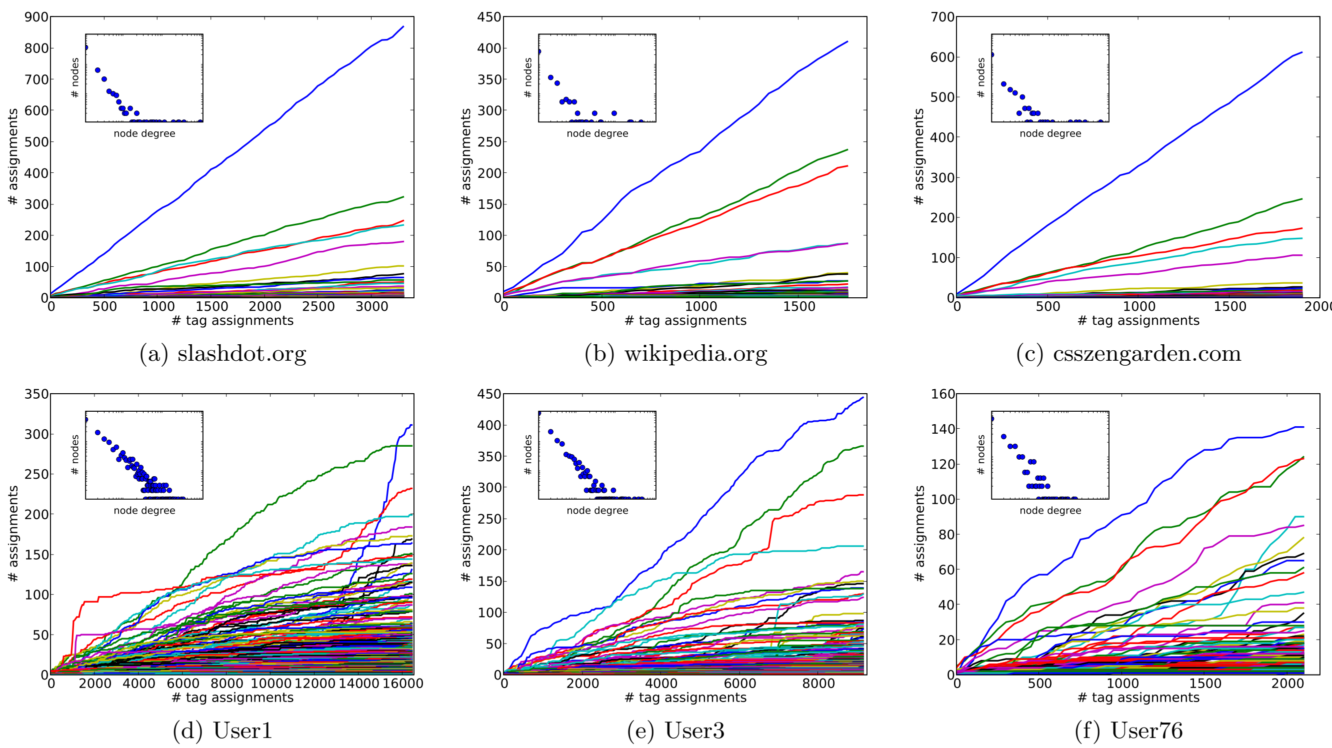


Figure 2: Tag distribution over time for three Delicious URLs and users. Each line corresponds to the frequency of one tag w.r.t the overall number of tag assignments (TAS). The small plots show the corresponding node degree distributions (*log-log-scale*). Resources develop a characteristic and stable tag spectrum strongly dominated by very few tags. User vocabularies, on the other hand, react more dynamically to shifting user's interests or tagging behavior. The node degree distributions of personomies exhibit power law characteristics, whereas the contrast between very frequent and very rare item tags even exceeds an expected power law divergence.

fact that though users agree on some characteristic resource labels – an agreement likely to be supported by existing tag recommenders – they additionally tend to assign their own sets of category tags. The diversity of these tags results in the observed long-tail distributions of the node degrees.

An exploration of the 500 most popular URLs in our Delicious corpus unveils stable tag distributions for all items, and we believe that this stability is a natural consequence of collaborative tagging. These scale-free spectra allow us to infer a folksonomy opinion about resources, even if these have been bookmarked by only a subset of users. This justifies our idea of translating to the community vocabulary, although we only deal with incomplete data. The stability of the community opinion is also important from a practical perspective, since we do not need to retrain our translational mappings whenever resources receive additional bookmarks.

A4: Personomies are stable in the short-term.

Our translation approach tacitly assumes that the user models, which have been trained on past data, are valid for the prediction of future user behavior. This makes it essential to look at how fast user interests and tagging strategies, that both manifest within personomies, actually change. Juxtaposing the personomies of Delicious users, we encounter more diversity then we did for item vocabularies. However, the majority of personomies develops similar to the examples that are displayed in the second row of Figure 2. Not unexpectedly, we observe that personomies are much less static than resource labels and reflect interest shifts or changing labeling strategies. User tag models, even though representative in the short term, will therefore require constant updates. Furthermore, comparing the tag distributions of URLs and users unveils that personomies tend to be less dominated by very frequent tags than are resources. Instead, the distributions of node degrees often look more similar to power law distributions which corresponds to linear relations on a *log-log-scale*. The deviating distribution patterns between personomy and item tags render it less efficient to directly compute user item similarities, e.g. by calculating the overlap between personomies and item vocalbularies [23]. Our approach overcomes these problems by mapping user tags to item derived vocabularies before calculating similarities.

Another way to investigate the stability of personomies is to look at the number of newly assigned tags that did not occur in the vocabulary before. Figure 1(c) illustrates that the percentage of new tags covered by a user's personomy constantly increases as the personomy grows. The first 10 tags a user assigns reoccur in more than 50 percent of the following labels which is a further indicator of the stability of user interests in the short term. The results of Figure 1(c) are also significant for the tag recommendation scenario, since they represent the maximum recall rates purely item or user vocabulary based recommenders can achieve given the tag assignments in the training data.

The assumptions explained in this section form the basic building blocks of our translation approach which we describe in the following section.

5. TRANSLATING PERSONOMIES

5.1 Notation

Following common notation conventions we use bold calligraphic letters for tensors ($\boldsymbol{\mathcal{Y}}$), bold capital letters for matrices ($\mathbf{A}$), and bold lower case letters for vectors ($\mathbf{u}$). Sets and set elements are represented by non-bold letters, e.g. I and i. Capital letter subscripts of tensors and matrices denote the sets corresponding to each mode, such that the matrix $\mathbf{A}_{IT}$ represents a $|I|$-by-$|T|$ matrix with rows and columns build from the sets I and T respectively. Lower case letter subscripts represent the reduction to only one entry of the corresponding set, e.g. $\mathbf{A}_{it}$ is the entry of $\mathbf{A}_{IT}$ that resides in row i and column t. Analogously, we will use subscripts for sets to indicate the members of the original set that fall into the context represented by the subscript, e.g. I_u represents all items a user u has bookmarked.

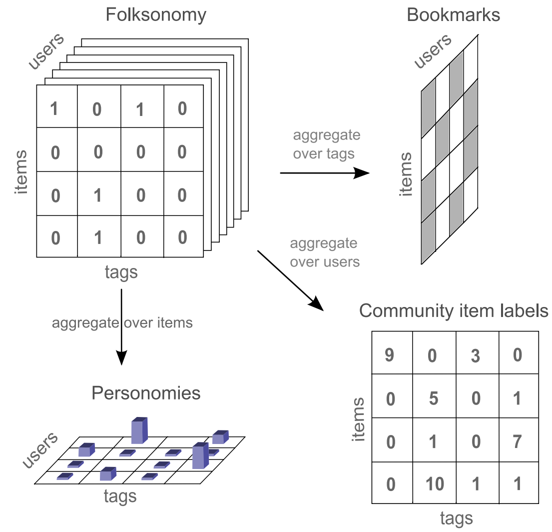


Figure 3: A folksonomy as a tensor and corresponding matrices after aggregating over one mode.

5.2 Formalization

A folksonomy can be represented by a tuple $F := (I, T, U, Y)$, where I, T, U are finite sets of items, tags, and users, and $Y \subseteq I \times T \times U$ is a ternary relation, whose elements are called *tag assignments* [11]. Bookmarks can then be formalized as the sets of tag assignments that share the same user item entities $B = \{(i, u)|\exists t : (i, t, u) \in Y\}$. The tag vocabulary of a user, her personomy, is given as $T_u = \{t|\exists i : (i, t, u) \in Y\}$, the tag vocabulary of an item as $T_i = \{t|\exists u : (i, t, u) \in Y\}$, and the tags of a bookmark as $T_{iu} = \{t|(i, t, u) \in Y\}$.

A natural way to represent the ternary relations Y that build a folksonomy are tensors as shown in Figure 3. We can construct the folksonomy tensor $\boldsymbol{\mathcal{Y}}_{ITU} \in \mathbb{R}^{|I|\times|T|\times|U|}$ that represents Y as

$$\boldsymbol{\mathcal{Y}}_{itu} = \begin{cases} 1, & \text{if } (i, t, u) \in Y \\ 0, & \text{otherwise.} \end{cases}$$

Modeling folksonomies as tensors allows us to apply tensor calculus to derive substructures of Y. For example, the co-occurrence matrix $\mathbf{A}_{IT}$ of items and tags, which we will later require for our translation model, can be calculated by aggregating the tags assigned to each item over all users. This is equivalent to

$$\mathbf{A}_{IT} = \boldsymbol{\mathcal{Y}}_{ITU} \,\overline{\times}_U\, \mathbf{u}^1, \tag{1}$$

where $\overline{\times}_X$ is the vector product with a tensor in the mode represented by X, and $\mathbf{u}^1$ is a vector of size $|U|$ of all ones. Each entry $\mathbf{A}_{it}$ of the matrix $\mathbf{A}_{IT}$ hence denotes the number of users that have labeled i with t. It is likewise possible to aggregate along other modes as shown in Figure 3. Using the vector product, we can also derive the tags a user assigned to an item as

$$\mathbf{t}(i,u) = \boldsymbol{\mathcal{Y}}_{ITU} \,\overline{\times}_I\, \mathbf{i} \,\overline{\times}_U\, \mathbf{u}, \tag{2}$$

where $\mathbf{i}$ and $\mathbf{u}$ are vectors of all zeros except for the indices that correspond to the user item pair and which are set to 1. Notation-wise, this is equivalent to writing $\boldsymbol{\mathcal{Y}}_{iTu}$.

5.3 Calculating the translation tensor

We assume that each user has a distinctive vocabulary of tags that can be translated to the folksonomy vocabulary by looking at tag co-occurrences within the shared item space. Employing tensor calculus, we can construct the tensor $\boldsymbol{\mathcal{T}}_{T\tilde{T}U}$ that contains these (item-normalized) tag co-occurrence counts for each user as

$$\boldsymbol{\mathcal{T}}_{T\tilde{T}U} = \boldsymbol{\mathcal{Y}}_{ITU} \times_I \mathbf{A'}_{TI}, \tag{3}$$

where the T and $\tilde{T}$ modes of the translation tensor represent the folksonomy and personomy tag dimensions, $\mathbf{A'}_{TI}$ is the transposed and item-normalized stochastic version of the global item tag co-occurrence matrix from Equation 1, and $\times_X$ denotes the matrix product with a tensor in the mode corresponding to the set X. The value of each element in $\boldsymbol{\mathcal{T}}_{T\tilde{T}U}$ is thus equal to

$$\boldsymbol{\mathcal{T}}_{t\tilde{t}u} = \sum_{i \in I} \boldsymbol{\mathcal{Y}}_{i\tilde{t}u} \mathbf{A'}_{ti}. \tag{4}$$

Since $\boldsymbol{\mathcal{Y}}_{i\tilde{t}u}$ will be 1 if $(i, \tilde{t}, u) \in Y$ and 0 otherwise, this can also be written as

$$\boldsymbol{\mathcal{T}}_{t\tilde{t}u} = \sum_{i \in I_{u\tilde{t}}} \mathbf{A'}_{ti}. \tag{5}$$

Each tag $\tilde{t}$ from the personomy of a user u is thus represented by a tag vector $\boldsymbol{\mathcal{T}}_{T\tilde{t}u}$ that aggregates the tag distributions of all items that user u has labeled with $\tilde{t}$. One important aspect of our model is that it considers the user as a part of the folksonomy she is mapped to. This allows us to derive primitive mappings even in cases where a user does not share any items with other users or where the shared items are sparsely tagged.

Based on $\boldsymbol{\mathcal{T}}_{T\tilde{T}U}$, tags from a user's personomy represented by $\tilde{\mathbf{t}}$ can now be translated to the folksonomy vocabulary by simple vector multiplication:

$$\mathbf{t}(\tilde{\mathbf{t}}, u) = \frac{\boldsymbol{\mathcal{T}}_{T\tilde{T}u} \,\overline{\times}_{\tilde{T}}\, \tilde{\mathbf{t}}}{|\boldsymbol{\mathcal{T}}_{T\tilde{T}u} \,\overline{\times}_{\tilde{T}}\, \tilde{\mathbf{t}}|}, \tag{6}$$

where $|.|$ represents the sum of all vector elements. Each value of $\mathbf{t}(\tilde{\mathbf{t}}, u)$ gives us a weight for the user meaning t when saying $\tilde{\mathbf{t}}$. Analogously, we can translate from the global tag language to a user's personomy:

$$\tilde{\mathbf{t}}(\mathbf{t}, u) = \frac{\boldsymbol{\mathcal{T}}_{T\tilde{T}u} \,\overline{\times}_{T}\, \mathbf{t}}{|\boldsymbol{\mathcal{T}}_{T\tilde{T}u} \,\overline{\times}_{T}\, \mathbf{t}|}. \tag{7}$$

We will later use the mapping from Equation 7 to identify the most likely user tags for an item given the tags previously assigned by the community. The translation of user tags to the global tag vocabulary from Equation 6, on the other hand, will help us to infer the meaning of a query tag in the search scenario.

6. SCENARIO 1: TAG RECOMMENDATION

Tag recommendation within a folksonomy is the task of predicting $\boldsymbol{\mathcal{Y}}_{iTu}$ for a previously unseen bookmark, i.e. $(i,u) \notin B$.

6.1 The tag recommender

We approach the problem of tag recommendation by translating the tag spectrum of the item i to the personomy of user u. The weights $\hat{\mathbf{t}}(i,u)$ for each personomy tag are calculated using Equation 7 as

$$\tilde{\mathbf{t}}(i,u) = \frac{\boldsymbol{\mathcal{T}}_{T\tilde{T}u} \,\overline{\times}_{T}\, \mathbf{A'}_{iT}}{|\boldsymbol{\mathcal{T}}_{T\tilde{T}u} \,\overline{\times}_{T}\, \mathbf{A'}_{iT}|}, \tag{8}$$

where $\mathbf{A'}_{iT} = \mathbf{A'}_{IT} \overline{\times}_I \mathbf{i}$ denotes the item's community tag distribution. Based on $\tilde{\mathbf{t}}(i,u)$ we can now predict the personomy tags a user will assign to a new item. However, $\tilde{\mathbf{t}}(i,u)$ only contains weights for tags within the user's personomy, whereas all tags the user did not assign yet will have zero weights. Recommending tags solely based on $\tilde{\mathbf{t}}(i,u)$ would thus disregard the previously noted fact that many tags are item specific (see also Figure 1(c)). We therefore extend our model to include item tags to

$$\hat{\mathbf{t}}(i,u) = \alpha\, \tilde{\mathbf{t}}(i,u) + (1-\alpha)\, \mathbf{t}(i), \tag{9}$$

with $\mathbf{t}(i) = \mathbf{A'}_{iT}$. Tags are thus recommended from a mixture of personomy and item tags, where the parameter $0 \leq \alpha \leq 1$ controls the mixture ratio. For each user item combination, we then recommend the N tags with the highest value in $\hat{\mathbf{t}}(i,u)$.

6.2 Experiments

We run our evaluations on p-core versions of the original Delicious and Bibsonomy datasets with p set to 5. From both datasets we randomly select 10 percent of all bookmarks for testing and keep the rest for training. For each test bookmark, we then predict a ranked list of tags based on $\hat{\mathbf{t}}(i,u)$ and calculate the achieved *Precision*, *Recall* and the corresponding F_1 measures at different ranks N. We report results averaged over all bookmarks and 10 test runs. The best value of the α parameter is determined using a *brute-force* approach. We choose the α value that performs best in terms of maximal reached F_1 measure.

6.2.1 Other recommenders

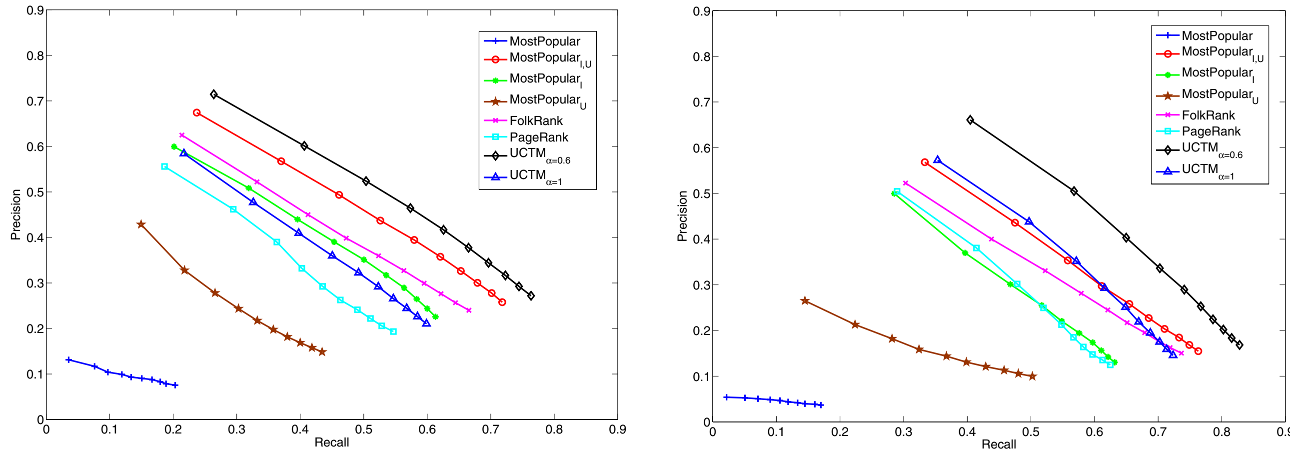
We compare the results achieved by our user-centric tag model (UCTM) approach to a variety of baseline recommenders.

MostPopular This recommender weights tags by their global occurrence in the training data.

MostPopular$_I$, MostPopular$_U$ These recommenders weight tags by their co-occurrence with the given item or user.

MostPopular$_{I,U}$ This recommender weights tags by a mixture of their item and user co-occurrence distributions

$$\hat{\mathbf{t}}(i,u) = \beta \frac{\mathbf{t}(i)}{|\mathbf{t}(i)|} + (1-\beta) \frac{\mathbf{t}(u)}{|\mathbf{t}(u)|},$$



(a) Bibsonomy (b) Delicious

Figure 4: Precision/Recall curves for various tag recommenders.

with $0 \leq \beta \leq 1$. A similar recommender was reported to perform reasonably well in [12].

Adapted PageRank Hotho et al. [11] propose an extension of the well-known PageRank algorithm [4] to folksonomies. Instead of the Web graph, they merge all node sets to a set $V = I \cup T \cup U$. They then construct a $|V| \times |V|$ weighted adjacency matrix $\mathbf{A}$, where each entry corresponds to the number of times two nodes co-occur in Y. The weight update process is defined as $\mathbf{w} \leftarrow d\mathbf{A}\mathbf{w}+(1-d)\mathbf{p}$, where $\mathbf{p}$ is a preference vector, and d regulates the impact of the preference vector. As proposed by [12], we set entries of $\mathbf{p}$ to 1 except for the entries that correspond to the input item and user and that are set to $|I|$ and $|U|$ respectively. d is set to 0.7. We stop the weight diffusion process after 10 steps.

FolkRank The fact that edges within folksonomies are undirected results in a suboptimal accuracy of the Adapted PageRank algorithm [11, 12]. As a consequence, Hotho et al. [11] propose a differential approach $\mathbf{w} = \mathbf{w}^1 - \mathbf{w}^0$, where $\mathbf{w}^0$ is the global weighting obtained from the Adapted PageRank algorithm when $\mathbf{p} = \mathbf{1}$, and $\mathbf{w}^1$ is the result vector with preference, as described above. This differential approach is known as the FolkRank algorithm.

The optimal value for the parameter β of the $MostPopular_{I,U}$ model is again determined by a *brute-force* approach. For all recommenders we always suggest the N tags with the highest weight.

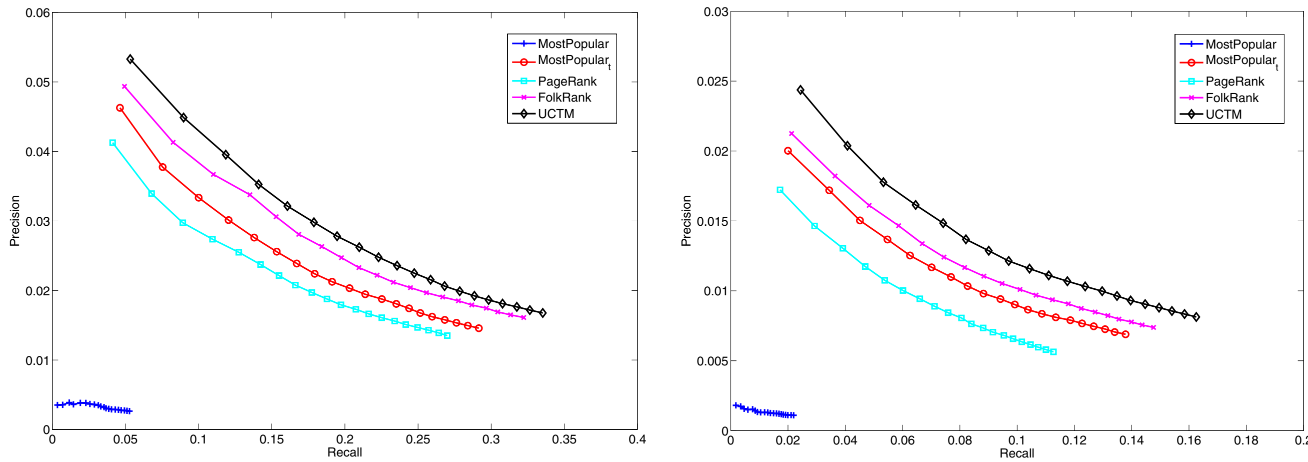
6.3 Results

Figure 4 presents the *Precision* and *Recall* values of all recommenders when suggesting up to 10 tags. Independent of the dataset, we observe very similar results, with our user-centric tag model ($UCTM$) performing best. The importance of tag translation is highlighted when comparing the results of the $MostPopular_U$ with the $UCTM_{\alpha=1}$ recommender. Both models only return the tags a user has already used before. However, selecting tags based on the (translated) community opinion about an item is much more accurate, since it respects the item's nature. The comparably good result of the $MostPopular_I$ recommender is most likely a consequence of the previously discussed very high density of item tags, which simplifies the prediction of popular tags. However, combining item and (translated) user tags still yields large performance gains, and a $MostPopular_{I,U}$ recommender with β set to 0.5 produces reasonable good results on both datasets, outperforming even more advanced approaches such as the $FolkRank$ method.

The best recommendations on both datasets are given by a $UCTM$ recommender with a mixture ratio of $\alpha = 0.6$. This recommender reaches a maximal F_1 measure of 0.514 on Bibsonomy and 0.533 on Delicious, performing significantly better than the $MostPopular_{I,U}$ (0.478/0.448) and the FolkRank (0.430/0.413) methods. The performance gain of our approach on the Delicious dataset is probably a consequence of larger personomy sizes that result in more robust translational models. The very good performance of our $UCTM$ approach underlines the potential of tag translation for the task of tag recommendation. This potential is even higher if we consider that our method of fusing translated user and item tag models in Equation 9 is rather basic and still leaves much space for improvements.

7. SCENARIO 2: SOCIAL SEARCH

In the social search scenario we consider a user who wants to find content related to one of the tags from her personomy. Search scenarios will generally require that the user guesses the tags other users might have assigned to relevant content. Instead, we enable the user to search based on her individual vocabulary and translate the user query to the global vocabulary. The proposed social search service thus supports users in discovering new content with respect to the topics they have previously shown interest in. For instance, *User237* from Table 2 may search for publications using the tag "diplomathesis" and expecting results similar to the documents he previously labeled with this tag. This allows her to discover content related to her own work of which she was previously unaware of. The challenge in this scenario is to identify the search intention that a user asso-



(a) Bibsonomy (b) Delicious

Figure 5: Precision/Recall curves for the social search task.

ciates with the given tag. From a technical perspective, we are required to predict $\boldsymbol{\mathcal{Y}}_{Itu}$, where $(i, u) \notin B$ and $t \in T_u$.

7.1 Translating query tags

We divide the search process into two steps: The translation of the query tag to the global vocabulary and the actual item retrieval. Translating a single user tag $\tilde{t}$ only requires to look at it's previous co-occurrence with other tags represented by the vector $\boldsymbol{\mathcal{T}}_{T\tilde{t}u}$ within the translation tensor $\boldsymbol{\mathcal{T}}_{T\tilde{T}U}$. In the second step, we then identify the items which generally fall together with these community tags calculating the weight vector $\widehat{\mathbf{i}}(\tilde{t}, u)$ as

$$\widehat{\mathbf{i}}(\tilde{t}, u) = \frac{\boldsymbol{\mathcal{T}}_{T\tilde{t}u} \ \times_T \ \mathbf{A^*}_{IT}}{|\boldsymbol{\mathcal{T}}_{T\tilde{t}u} \ \times_T \ \mathbf{A^*}_{IT}|}, \tag{10}$$

where $\mathbf{A^*}_{IT}$ is the tag-normalized stochastic version of the matrix $\mathbf{A}_{IT}$. In this work, we only consider single-tag search queries. However, our model can be easily extended to multi-tag queries by replacing $\boldsymbol{\mathcal{T}}_{T\tilde{t}u}$ with $\boldsymbol{\mathcal{T}}_{T\tilde{T}u} \overline{\times}_{\tilde{T}} \ \tilde{\mathbf{t}}$, where $\tilde{\mathbf{t}}$ is the tag vector corresponding to the query. We set all values of $\widehat{\mathbf{i}}(\tilde{t}, u)$ to 0 where $(i, u) \in B$, since items are bookmarked only once per user.

7.2 Experiments

We apply the same test setup as in the tag recommendation scenario. For each bookmark of the test set we now randomly select a tag to simulate a query and try to predict the item the user had labeled with this tag. To avoid item specific tags with little expressiveness for user interests, we require that each query tag was applied at least 5 times by the user in the training data. For each user tag combination we then recommend the N items with highest weight in $\widehat{\mathbf{i}}(\tilde{t}, u)$. As for the tag recommendation scenario, we test our approach against a variety of other methods, including the previously described **Adapted PageRank** and **FolkRank** algorithms. Furthermore, we report results for a **MostPopular** item recommender as well as for a tag-aware item recommender (**MostPopular**$_T$) that ranks items according to their previous co-occurrence with the query tag, but does not perform any personalization.

7.3 Results

The *Recall* and *Precision* values for the first 20 ranks and for all search methods are displayed in Figure 5. Both plots demonstrate a superior result quality for our translation approach ($UCTM$). The query tag translation improves the *Recall@10* by about 15 percent on Bibsonomy and 18 percent on Delicious compared to the $MostPopular_T$ baseline method. Our translational model also performs better than the $FolkRank$ algorithm, yielding *Recall@10* improvements of 7.5 percent and 10 percent. This becomes even more meaningful if we consider that the $FolkRank$ algorithm incorporates collaborative filtering information, e.g. information about users with similar bookmarking patterns, whereas our approach personalizes solely by trying to understand the user's search intention. Similar to the tag recommendation task, we find the *AdaptedPageRank* algorithm to perform worse than simple baseline methods.

The tag recommendation scenario has shown that global tag distributions can be translated to local, user-level distributions. The good results of our approach in the social search scenario, on the other hand, show that this translation can be performed in both directions, resulting in better user models and an improved awareness of user intentions.

The prediction task in this search setup reoccurs in other application scenarios, such as the topic-aware recommendation of items or the topic-aware detection of new content. We believe that our translation approach can also help to identify user interests in these scenarios.

8. CONCLUSIONS

This work investigated the nature of tagging within folksonomies from a user-centric perspective. Our studies on two real-world social bookmarking services revealed that users who tag for content categorization develop distinct tag vocabularies over time. This heterogeneity vanishes when the tags of many users are aggregated, resulting in characteristic tag distributions for resources. However, the personal aspect of tagging cannot be ignored in user-centric scenarios, such as tag recommendation or tag-based search.

We presented a novel approach to tag translation, which maps user tags to the global folksonomy vocabulary using the labeled resources as intermediates. Based on these mappings, we were able to infer the meaning of a user tag and to accurately predict which tags a user will assign to new content. We evaluated the applicability of our approach in two relevant use cases: tag recommendation and tag-based social search. Our experiments reveal that tag translation improves prediction accuracy in both scenarios.

This paper focused on tagging in broad folksonomies where tags can be assigned by the entire community and users mainly tag for content organization. In future work we plan to extend our investigations to narrow folksonomies that exhibit flat tag distributions and display different tagging motivations. Narrow folksonomies would especially benefit from tag translation, since here user-level peculiarities in tagging are not compensated by aggregative effects which hinders the tag-based indexing of resources. Further, we want to obtain a better understanding of how shifts in user interests are reflected in a user's tagging behavior. Understanding this relationship will allow for the design of more accurate user models, which in turn can help improve the quality of many services.

9. ACKNOWLEDGMENTS

The authors would like to express their gratitude to Leonhard Hennig, Jeff Mounzer, Alan Said, Jan Hendrik Clausen, Thomas Karbe, and Jérôme Kunegis for valuable suggestions on early versions of this paper.

10. REFERENCES

[1] M. Ames and M. Naaman. Why we tag: motivations for annotation in mobile and online media. In *CHI '07*.

[2] C.-m. Au Yeung, N. Gibbins, and N. Shadbolt. Contextualising tags in collaborative tagging systems. In *HYPERTEXT '09*.

[3] K. Bischoff, C. S. Firan, W. Nejdl, and R. Paiu. Can all tags be used for search? In *CIKM '08*.

[4] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Computer Networks and ISDN Systems*, 1998.

[5] C. Cattuto, D. Benz, A. Hotho, and G. Stumme. Semantic grounding of tag relatedness in social bookmarking systems. In *ISWC '08*.

[6] S. A. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. *J. of Information Science*, 32(2), 2006.

[7] Z. Guan, J. Bu, Q. Mei, C. Chen, and C. Wang. Personalized tag recommendation using graph-based ranking on multi-type interrelated objects. In *SIGIR '09*.

[8] H. Halpin, V. Robu, and H. Shepherd. The complex dynamics of collaborative tagging. In *WWW '07*.

[9] P. Heymann, G. Koutrika, and H. Garcia-Molina. Can social bookmarking improve web search? In *WSDM '08*.

[10] P. Heymann, D. Ramage, and H. Garcia-Molina. Social tag prediction. In *SIGIR '08*.

[11] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In *ESWC '06*.

[12] R. Jäschke, L. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme. Tag recommendations in folksonomies. In *Lernen - Wissensentdeckung - Adaptivität (LWA '07) Workshop*.

[13] B. Krause, A. Hotho, and G. Stumme. A comparison of social bookmarking with traditional search. In *ECIR '08*.

[14] X. Li, L. Guo, and Y. E. Zhao. Tag-based social interest discovery. In *WWW '08*.

[15] M. Lipczak. Tag recommendation for folksonomies oriented towards individual users. In *Discovery Challenge* (ECML-PKDD '08).

[16] 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*.

[17] C. Marlow, M. Naaman, D. Boyd, and M. Davis. Ht06, tagging paper, taxonomy, flickr, academic article, to read. In *HYPERTEXT '06*.

[18] P. Mika. Ontologies are us: A unified model of social networks and semantics. In *ISWC 2005*.

[19] T.-G. Noh, S.-B. Park, H.-G. Yoon, S.-J. Lee, and S.-Y. Park. An automatic translation of tags for multimedia contents using folksonomy networks. In *SIGIR '09*.

[20] S. Oldenburg, M. Garbe, and C. Cap. Similarity cross-analysis of tag / co-tag spaces in social classification systems. In *Search in Social Media (SSM) Workshop* (CIKM '08).

[21] D. Ramage, P. Heymann, C. D. Manning, and H. Garcia-Molina. Clustering the tagged web. In *WSDM '09*.

[22] S. Rendle, L. Balby Marinho, A. Nanopoulos, and L. Schmidt-Thieme. Learning optimal ranking with tensor factorization for tag recommendation. In *KDD '09*.

[23] A. Shepitsen, J. Gemmell, B. Mobasher, and R. Burke. Personalized recommendation in social tagging systems using hierarchical clustering. In *RecSys '08*.

[24] Y. Song, Z. Zhuang, H. Li, Q. Zhao, J. Li, W.-C. Lee, and C. L. Giles. Real-time automatic tag recommendation. In *SIGIR '08*.

[25] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos. Tag recommendations based on tensor dimensionality reduction. In *RecSys '08*.

[26] T. V. Wal. Explaining and showing broad and narrow folksonomies, 2005. http://www.vanderwal.net.

[27] T. V. Wal. Folksonomy coinage and definition, 2007. http://www.vanderwal.net/folksonomy.html.

[28] R. Wetzker, W. Umbrath, and A. Said. A hybrid approach to item recommendation in folksonomies. In *Exploiting Semantic Annotations in IR (ESAIR) Workshop* (WSDM '09).

[29] R. Wetzker, C. Zimmermann, and C. Bauckhage. Analyzing social bookmarking systems: A del.icio.us cookbook. In *Mining Social Data (MSoDa) Workshop* (ECAI '08).

[30] X. Wu, L. Zhang, and Y. Yu. Exploring social annotations for the semantic web. In *WWW '06*.

[31] S. Xu, S. Bao, B. Fei, Z. Su, and Y. Yu. Exploring folksonomy for personalized search. In *SIGIR '08*.

[32] V. Zanardi and L. Capra. Social ranking: uncovering relevant content using tag-based recommender systems. In *RecSys '08*.