# Forgetting the Words but Remembering the Meaning: Modeling Forgetting in a Verbal and Semantic Tag Recommender

Dominik Kowald[1,2], Paul Seitlinger[2], Simone Kopeinik[2], Tobias Ley[3], and Christoph Trattner[4]

[1] Know-Center, Graz University of Technology, Graz, Austria
dkowald@know-center.at
[2] Knowledge Technologies Institute, Graz University of Technology, Graz, Austria
{paul.seitlinger,simone.kopeinik}@tugraz.at
[3] Institute of Informatics, Tallin University, Tallinn, Estonia
tley@tlu.ee
[4] Norwegian University of Science and Technology, Trondheim, Norway
chritrat@idi.ntnu.no

**Abstract.** We assume that recommender systems are more successful, when they are based on a thorough understanding of how people process information. In the current paper we test this assumption in the context of social tagging systems. Cognitive research on how people assign tags has shown that they draw on two interconnected levels of knowledge in their memory: on a conceptual level of semantic fields or LDA topics, and on a lexical level that turns patterns on the semantic level into words. Another strand of tagging research reveals a strong impact of time-dependent forgetting on users' tag choices, such that recently used tags have a higher probability being reused than "older" tags. In this paper, we align both strands by implementing a computational theory of human memory that integrates the two-level conception and the process of forgetting in form of a tag recommender. Furthermore, we test the approach in three large-scale social tagging datasets that are drawn from BibSonomy, CiteULike and Flickr.

As expected, our results reveal a selective effect of time: forgetting is much more pronounced on the lexical level of tags. Second, an extensive evaluation based on this observation shows that a tag recommender interconnecting the semantic and lexical level based on a theory of human categorization and integrating time-dependent forgetting on the lexical level results in high accuracy predictions and outperforms other well-established algorithms, such as Collaborative Filtering, Pairwise Interaction Tensor Factorization, FolkRank and two alternative time-dependent approaches. We conclude that tag recommenders will benefit from going beyond the manifest level of word co-occurrences, and from including forgetting processes on the lexical level.

**Keywords:** personalized tag recommendations; time-dependent recommender systems; Latent Dirichlet Allocation; LDA; human categorization; human memory model; BibSonomy; CiteULike; Flickr

## 1 Introduction

Many interactive systems are designed to mimic human behavior and thinking. A telling example for this are intelligent tutoring systems, which make inferences similar to teachers when drawing on knowledge of learning domains, knowledge about the learners and knowledge about effective teaching strategies. When looking at recommender systems, Collaborative Filtering approaches use information about socially related individuals to recommend items, much in the same way as humans are influenced by related peers when making choices. An implicit assumption behind this may be that interactive systems will perform better the closer they correspond to human behavior. This assumption seems to be reasonable as it is humans that interact with these systems, while these systems often also draw on data produced by humans (e.g., in the case of Collaborative Filtering). Therefore it can be assumed, that strategies that have evolved in humans over their individual or collective development form good models for interactive systems. However, the assumption that an interactive system will perform better the closer it mimics human behavior has not often been tested directly.

In the current paper, we investigate this assumption in the context of a tag recommender algorithm that borrows its basic architecture from MINERVA2 ([1], see also [2]), a computational theory of human categorization. We draw on research that has explored how human memory is used in a dynamic and adaptive fashion to understand new information encountered in the environment. Sensemaking happens by dynamically forming ad-hoc categories that relate the new information with knowledge stored in the semantic memory (e.g., [3]). For instance, when reading an article about "personalized recommendations", a novice has to figure out meaningful connections between previously distinct topics such as "cognition" and "information retrieval" and hence, has to start developing an ad-hoc category about common features of both of them. When using a social tagging system in such a situation, people apply labels to their own resources which to some extent externalizes this process of spontaneously generating ad-hoc categories [4]. Usually, a user describes a particular bookmark by a combination of about three to five tags verbalizing and associating aspects of different topics (e.g., "memory","retrieval", "recommendations", "collaborative filtering").

In previous work, we have shown that this behavior can be well described by differentiating between two separate forms of information processing. In human memory we find a semantic process that generates and retrieves topics or gist traces, and a verbal process that generates verbatim word forms to describe the topics [5]. In this paper we improve this model emphasizing on another fundamental principle of human cognition. According to Polyn et al. [6], memory traces including recently activated features contribute more strongly to retrieval than traces including features that have not been activated for a longer period of time. This relationship provides a natural account of what is called the recency effect in memory psychology (e.g., [7]). Obviously, things that happened a longer time ago tend to be forgotten and influence our current behavior less than things that have happened recently.

The purpose of this paper is twofold. First, we study the interaction between the effect of recency and the level of knowledge representation in human memory (semantic vs. verbal) within a social tagging system. In particular, we raise the question whether the impact of recency interacts with the level of knowledge representation, i.e., whether a time-dependent shift in the use of topics can be dissociated from a time-dependent shift in the use of particular tags. The second aim is to investigate to which extend our tag recommender based on MINERVA2 can be improved by integrating a time-dependent forgetting process. We also determine the performance of this recommender compared to other well-established tag recommender algorithms (e.g., Collaborative Filtering, FolkRank and Pairwise Interaction Tensor Factorization), as well as two alternative time-dependent approaches called GIRPTM [8] and BLL+C [9] (based on the ACT-R theory of human memory [10]). Hence, we raise the following two research questions:

- *RQ1*: Is there a difference between the time-dependent shift in the use of topics and the time-dependent shift in the use of particular tags?
- *RQ2*: Can a time-dependent forgetting process be integrated into a tag-recommender to create an efficient algorithm in comparison to the state-of-the-art?

The remainder of this paper tackles this two research questions and is organized as follows: We begin with discussing related work in the field of tag recommender in Section 2. Next, we review some of the work concerning recency in memory research and its current use in social tagging in Section 3 (first research question). Then we describe our approach and the experimental setup of our extensive evaluation in Sections 4 and 5. Section 6 presents the results of this evaluation in terms of recommender quality (second research question). We conclude the paper by discussing our findings and future work in Section 7.

## 2 Related Work

Tagging as an important feature of the Social Web, has demonstrated to improve search considerably [11, 12] and has supported the users with a simple tool to collaboratively organize and annotate content [13]. However, despite the potential advantages of tag usage, people do not tend to provide tags thoroughly or regularly. Thus, from an applied perspective, one important purpose of tag recommendations is to increase user's motivation to provide appropriate tags to their bookmarked resources.

In contrast to previously developed and typically data-driven tag recommender approaches, our research explores the suitability of psychologically sound memory processes to improve tag recommender approaches. Previously, in [5, 9] we presented two simple methods (= 3L and BLL+C) that aim to explain memory processes in social tagging systems. Based on our previous research and other incentives from related work, we introduce in this work a novel time-based tag

recommender algorithm (= $3LT_{tag}$) based on the MINERVA2 theory of human categorization [1, 2] that significantly outperforms popular state-of-the-art algorithms as well as BLL+C [9], an alternative time-based approach based on the ACT-R theory of human memory [10]. It models the activation of elements in a person's declarative memory by considering frequency and recency of a user's tagging history as well as semantic context.

To date, two tag-recommender approaches have been established: graph-based and content-based tag recommender systems [14], whereas in this work we focus on graph-based approaches. Prominent algorithms in this respect can be found for instance in the work of Hotho et al. [15] who introduced FolkRank (FR), which has established itself as the most prominent benchmarking tag recommender approach over the past few years. Further investigated, was the recommendation of tags to users in a personalized manner. In the scope of this research strand, Jäschke et al. [16] or Hamouda & Wanas [17] are well known to present a set of Collaborative Filtering (CF) approaches. Rendle et al. [18], Krestel et al. [19] or Rawashdeh et al. [20] more recently presented a factorization model (FM and PITF), a semantic model (based on LDA) or a link prediction model to recommend tags to users, respectively (see also Section 5.3).

Comparing these principles now with simple "most popular tags" approaches, we will notice a big disadvantage in their computational expense as well as in their lack of considering recent observations made in social tagging systems, such as the variation of the individual tagging behavior over time [21]. To that end, recent research has made first promising steps towards more accurate graph-based models that also account for the variable of time [22, 8].

However, although these time-dependent approaches have shown to outperform some of the current state-of-the-art tag recommender algorithms, all of them ignore well-established and long standing research from cognitive psychology on how humans process information. Therefore, we try to fill this gap by investigating tagging mechanisms that aim to mimic peoples' tagging behavior.

## 3 Recency in Memory and in the Use of Social Tagging

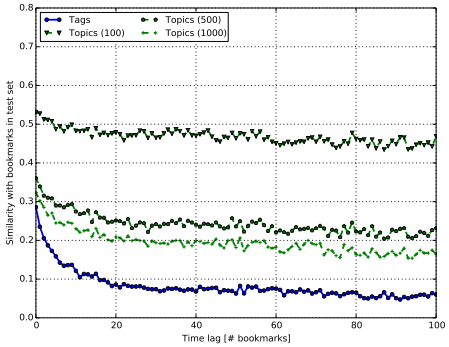
In previous work we have introduced 3Layers [5], which is a model for recommending tags that is inspired by cognitive-psychological research on categorizing and verbalizing objects (e.g., [4]) and is adapted in this work based on MINERVA2 in order to answer our two research questions. 3Layers consists of an input, a hidden and an output layer, where the hidden layer is built up by a semantic and an interconnected lexical matrix. The semantic matrix stores the topics of all bookmarks in the user's personomy[5], calculated with Latent Dirichlet Allocation (LDA) [19], while the lexical matrix stores the tags of those bookmarks. In a first step of calculation, the LDA topics of a new bookmark,

[5] We define a bookmark (also known as "post") as the set of tags a target user has assigned to a target resource at a specific time, and the personomy as a collection of all bookmarks of a user.

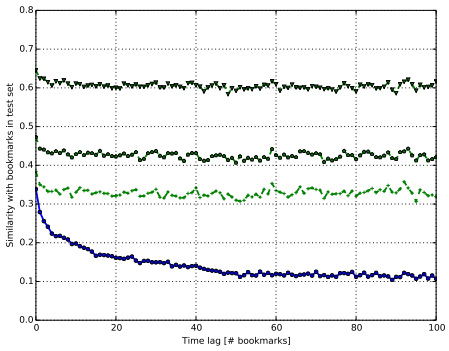
for which appropriate tags should be recommended, are represented in the input layer and compared with the semantic matrix of the hidden layer. In the course of this comparison, semantically relevant bookmarks of the user's personomy become activated. The resulting pattern of activation across the semantic matrix is then applied to the lexical matrix to further activate and recommend those tags that belong to relevant bookmarks. In a final step, the activation pattern across the lexical matrix is summarized on the output layer in form of a vector. This vector represents a tag distribution that can be used to predict a substantial amount of variance in the user's tagging behavior when creating a new bookmark.

We draw on Fuzzy Trace Theory (FTT; e.g., [23]) to make a prediction with respect to our first research question about a potentially differential impact of recency on semantic and lexical representations, i.e., on the usage of topics and tags, respectively. FTT differentiates between two distinct memory traces, a gist trace and a verbatim trace, which represent general semantic information of e.g., a read sentence and the sentence's exact wording, respectively. These two types of memory traces share properties with our distinction between a semantic and a lexical matrix (see also Section 4). While vectors of the semantic matrix provide a formal account of each bookmark's gist (its general semantic content), vectors of the lexical matrix correspond to a bookmark's verbatim trace (explicit verbal information in form of assigned tags). This distinction is also in line with Kintsch & Mangalath [24] who model gist traces of words by means of LDA topic vectors and explicit traces of words by means of word co-occurrence vectors. An empirically well-established assumption of FTT is that verbatim traces are much more prone to time-dependent forgetting than gist traces (e.g., [23]): while people tend to forget the exact wording, usually they can remember the gist of a sentence (or a bookmark). Taken together, we derived the hypothesis that a user's verbatim traces (vectors in the lexical matrix that encode the user's tags) are more strongly affected by time-dependent forgetting and therefore more variable over time than a user's gist traces (vectors in the semantic matrix that contain topics).

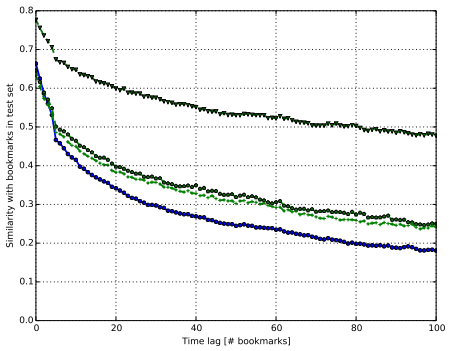
To test this hypothesis, we performed an empirical analysis in BibSonomy, CiteULike and Flickr (see Section 5.1). The topics for the resources of these datasets' bookmarks were calculated using Latent Dirichlet Allocation (LDA) [19] (see Section 4.2) based on 100, 500 and 1000 latent topics in order to cover different levels of topic specialization (these numbers of latent topics are also suggested by related work in the field [24, 25]). For each user we selected the most recent bookmark (i.e., the one from the test set with the most recent timestamp, see also Section 5.2) and described the bookmark by means of two vectors: one encoding the bookmark's LDA topic pattern (gist vector) and one encoding the tags assigned by the user (verbatim vector). Then, we searched for all the remaining bookmarks of the same user, described each of them by means of the two vectors and arranged them in a chronologically descending order. Next, we compared the gist and the verbatim vector of the most recent




(a) BibSonomy




(b) CiteULike




(c) Flickr

**Fig. 1.** Interaction between time-dependent forgetting and level of knowledge representation for BibSonomy, CiteULike and Flickr showing a more pronounced decline for tags than for topics (100, 500, 1000 LDA topics; first research question).

bookmark with the two corresponding vectors of all bookmarks in the user's past by means of the cosine similarity measure.

The obtained results are represented in the three diagrams of Figure 1, plotting the average cosine similarities over all users against the time lags (given in number of past bookmarks). For all three datasets we show these results for the last 100 bookmarks of tagging activity per user because in this range, there are enough users available for each past bookmark to calculate mean values reliably. The diagrams quite clearly reveal that – independent of the environment (BibSonomy, CiteULike or Flickr) – the similarity between the most recent bookmark and all other bookmarks decreases monotonically as a function of time lag. More importantly and as expected, the time-dependent decline is more strongly pronounced for the verbatim vectors (encoding tag assignments) in contrast to the gist vectors (encoding LDA topics). Furthermore, we can see that the more LDA topics we use, the more similar is the time-dependent decline of the two vectors (tags vs. topics) to each other.

## 4 Approach

In this section we introduce two novel time-dependent tag recommender algorithms which model the process of forgetting on a semantic and lexical layer in a time-depended manner. Moreover, we describe how we created the semantic features (i.e, topics) for the bookmarks in our datasets using *Latent Dirichlet Allocation* (LDA).

### 4.1 Tag-Recommender Algorithms

Due to our findings introduced within the previous section, we assume that the factor of time plays a more critical role on the lexical layer than on the semantic layer. The approaches implemented in this section are based on a preliminary recommender model called 3Layers (3L) that was introduced in our previous work [5].

Figure 2 schematically shows how 3Layers (3L) represents a user's personomy within the hidden layer, which interconnects a semantic matrix, $M_S$ ($l$ bookmarks $\cdot$ $n$ LDA topics matrix), and a lexical matrix, $M_L$ ($l$ bookmarks $\cdot$ $m$ tags matrix). Thus, each bookmark of the user is represented by two associated vectors; by a vector of LDA topics $S_{i,k}$ stored in $M_S$ and by a vector of tags $L_{i,j}$ stored in $M_L$. Similar to [2], we borrow a mechanism from MINERVA2, a computational theory of human categorization [1], to process the network constituted by the input, hidden and output layer. First, the LDA topics of the target resource to be tagged are represented on the input layer in form of a vector $P$ with $n$ features. Then, $P$ is used as a cue to activate each bookmark ($B_i$) in $M_S$ depending on the similarity ($Sim_i$) between both vectors, i.e., $P$ and $B_i$. Similar to [2], we estimate $Sim_i$ by calculating the cosine between the two vectors:

$$Sim_i = \frac{\sum_{k=1}^{n}(P_k \cdot S_{i,k})}{\sqrt{\sum_{k=1}^{n} P_k^2} \cdot \sqrt{\sum_{k=1}^{n} S_{i,k}^2}} \quad (1)$$

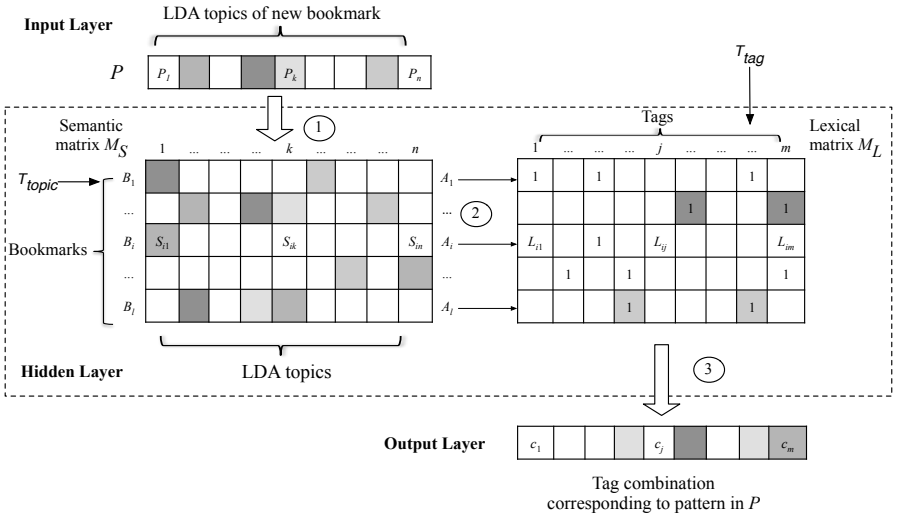



**Fig. 2.** Schematic illustration of 3L showing the connections between the semantic matrix ($M_S$) encoding the LDA topics and the lexical matrix ($M_L$) encoding the tags. Furthermore, $\mathrm{T}_{topic}$ and $\mathrm{T}_{tag}$ schematically demonstrate how the time component is integrated in case of $\mathrm{3LT}_{topic}$ and $\mathrm{3LT}_{tag}$, respectively.

If no topics are available for the target resource (i.e., $n = 0$), we set $Sim_i$ to 1 and thus, activate each bookmark with the same value. To transform the resulting similarity values into activation values ($A_i$) and to further reduce the influence of bookmarks with low similarities, $Sim_i$ is raised to the power of 3, i.e., $A_i = Sim_i^3$ (see also [1]). Next, these activation values are propagated to $M_L$ to activate tags that are associated with highly activated bookmarks on the semantic matrix $M_S$ (circled numbers 2 and 3 in Figure 2). This is computed by the following equation that yields an activation value $c_j$ for each of the $m$ tags on the output layer:

$$\underbrace{c_j = \sum_{i=1}^{l} (L_{i,j} \cdot A_i)}_{3L} \tag{2}$$

To finally compute $\mathrm{3LT}_{topic}$ and $\mathrm{3LT}_{tag}$, we integrate a time component on the level of topics (hereinafter called $\mathrm{T}_{topic}$) and on the level of tags ($\mathrm{T}_{tag}$), respectively. Both recency components are calculated by the following equation that is based on the base-level learning (BLL) equation [7]:

$$BLL(t) = ln((tmstp_{ref} - tmstp_t)^{-d}) \tag{3}$$

, where $tmstp_{ref}$ is the timestamp of the most recent bookmark of the user and $tmstp_t$ is the timestamp of the last occurrence of $t$, encoded as the topic

in the case of $\mathrm{T}_{topic}$ or as the tag in the case of $\mathrm{T}_{tag}$, in the user's bookmarks. The exponent $d$ accounts for the power-law of forgetting and was set to .5 as suggested by Anderson et al. [10]. While $3\mathrm{LT}_{topic}$ can be computed by using equation 4, $3\mathrm{LT}_{tag}$ can be computed by using equation 5:

$$\underbrace{c_j = \sum_{i=1}^{l}(L_{i,j} \cdot \sum_{k=1}^{n}(S_{i,k} \cdot BLL(k)) \cdot A_i)}_{3LT_{topic}} \tag{4}$$

$$\underbrace{c_j = \sum_{i=1}^{l}(L_{i,j} \cdot BLL(j) \cdot A_i)}_{3LT_{tag}} \tag{5}$$

As suggested in related work [26, 14, 9], we additionally consider tags that have been applied to the target resource by other users. This allows the recommendation of new tags, i.e., tags that have not been used by the target user before. We implement this by taking into account the most popular tags in the tag assignments of the resource $\mathrm{Y}_r$ ($\mathrm{MP}_r$, i.e., $\arg\max_{t \in T}(|Y_r|)$) [15]. Therefore, we have chosen $\mathrm{MP}_r$ over other methods like CF, as previous work [27, 28] shows that users in social tagging systems are more likely to imitate previously assigned tags by other users to a target resource. In order to combine $c_j$ with $\mathrm{MP}_r$, the following normalization method was used:

$$\|c_j\| = \frac{exp(c_j)}{\sum_{i=1}^{m} exp(c_i)} \tag{6}$$

Taken together, the list of recommended tags for a given user $u$ and resource $r$ is then calculated as

$$\widetilde{T}(u, r) = \arg\max_{j \in T}(\beta\|c_j\| + (1 - \beta)\||Y_r|\|) \tag{7}$$

, where $\beta$ is used to inversely weight the two components. The results presented in Section 6 were calculated using $\beta = .5$, thus, applying the same weight to both components.

### 4.2 Topic Generation via LDA

As outlined in Section 3, we used LDA to calculate the semantic features (i.e., topics) of the resources of the full datasets. LDA is a probability model that helps to find latent topics for documents where each topic is described by words in these documents [19]. This can be formalized as follows:

$$P(t_i|d) = \sum_{j=1}^{Z}(P(t_i|z_i = j) \cdot P(z_i = j|d)) \tag{8}$$

**Table 1.** Properties of the used dataset samples, where $|B|$ is the number of bookmarks, $|U|$ the number of users, $|R|$ the number of resources, $|T|$ the number of tags and $|TAS|$ the number of tag assignments.

| Dataset | $\|B\|$ | $\|U\|$ | $\|R\|$ | $\|T\|$ | $\|TAS\|$ |
|---|---|---|---|---|---|
| BibSonomy | 400,983 | 5,488 | 346,444 | 103,503 | 1,479,970 |
| CiteULike | 379,068 | 8,322 | 352,343 | 138,091 | 1,751,347 |
| Flickr | 864,679 | 9,590 | 864,679 | 127,599 | 3,552,540 |

Here $P(t_i|d)$ is the probability of the $i$th word for a document $d$ and $P(t_i|z_i = j)$ is the probability of $t_i$ within the topic $z_i$. $P(z_i = j|d)$ is the probability of using a word from topic $z_i$ in the document. The number of latent topics $Z$ is determined in advance and defines the level of granularity. We calculated the semantic features for our datasets with different amounts of LDA topics (100, 500 and 1000 - see also [24, 25]).

When using LDA in tagging environments, documents are resources which are described by tags. This means that based on the tag vectors of the resources (i.e., all the tags the users have assigned to the resource), resources in the bookmarks can also be represented with the topics identified by LDA. These topics were then used as features in the semantic matrix $M_S$. We implemented LDA with Gibbs sampling using the Java framework Mallet[6].

## 5 Experimental Setup

In this section we describe our experiment's datasets, evaluation methodology and the baseline algorithms in detail.

### 5.1 Datasets

To conduct our study, we used three well-known folksonomy datasets that are freely available for scientific purposes and thus, allow for reproducibility. In this respect, we utilized datasets from the social bookmark and publication sharing system BibSonomy[7] (2013-07-01), the reference management system CiteULike[8] (2013-03-10) and the image sharing platform Flickr[9] (2010-01-07) to evaluate our approach on both types of folksonomies, broad (BibSonomy and CiteULike; all users are allowed to annotate a particular resource) and narrow (Flickr; only the user who has uploaded a resource is allowed to tag it) ones [29]. We furthermore excluded all automatically generated tags from the datasets (e.g., *no-tag*, *bibtex-import*, etc.) and decapitalized all tags as suggested in related work (e.g., [18]). To reduce computational effort, we randomly selected 10% of CiteULike, and

[6] http://mallet.cs.umass.edu/topics.php
[7] http://www.kde.cs.uni-kassel.de/bibsonomy/dumps/
[8] http://www.citeulike.org/faq/data.adp
[9] http://www.tagora-project.eu/data/#flickrphotos

3% of Flickr user profiles (see also [30])[10]. We did not apply a $p$-core pruning to keep the original bookmarks of the users and thus, to prevent a biased evaluation [31]. The statistics of our used dataset samples can be found in Table 1.

### 5.2 Evaluation Methodology

To evaluate our tag recommender approaches, we split the three datasets into training and test sets based on a leave-one-out hold-out method as proposed in related work (e.g., [16]). Hence, for each user we selected her most recent bookmark (or post) in time and put it into the test set. The remaining bookmarks were then used for the training of the algorithms. This procedure is a promising simulation of a real-world environment, as it predicts a user's future tagging behavior based on tagging behavior in the past. Furthermore, it is a standard practice for evaluation of time-based recommender systems [32].

In order to quantify the recommender quality and to benchmark our recommender against other tag recommendation approaches, a set of well-known metrics in information retrieval and recommender systems were used [16, 14]:

**Recall (R)** is defined as the number of recommended tags that are relevant for the target user/resource divided by the total number of relevant tags [33]:

$$R@k = \frac{1}{|U|} \sum_{u \in U} \left(\frac{|t_u^k \cap T_u|}{|T_u|}\right) \tag{9}$$

, where $t_u^k$ denotes the top $k$ recommended tags and $T_u$ the list of relevant tags of a bookmark of user $u \in U$.

**Precision (P)** is calculated as the number of correctly recommended tags divided by the total number of recommended tags $|t_u^k|$ $(= k)$ [33]:

$$P@k = \frac{1}{|U|} \sum_{u \in U} \left(\frac{|t_u^k \cap T_u|}{|t_u^k|}\right) \tag{10}$$

**F1-score (F1)** is a combination of the recall and precision metrics and is calculated using the following equation [33]:

$$F1@k = \frac{1}{|U|} \sum_{u \in U} \left(2 \cdot \frac{P@k \cdot R@k}{P@k + R@k}\right) \tag{11}$$

**Mean reciprocal rank (MRR)** is a rank-dependent evaluation metric that is calculated as the sum of the reciprocal ranks (or positions) of all relevant tags in the list of recommended tags [20]:

$$MRR = \frac{1}{|U|} \sum_{u=1}^{|U|} \left(\frac{1}{|T_u|} \sum_{t \in T_u} \frac{1}{rank(t)}\right) \tag{12}$$

[10] **Note:** We used the same dataset samples as in our previous work [9], except for CiteULike, where we used a smaller sample for reasons of computational effort in respect to the calculation of the LDA topics.

This way, a recommender achieves a higher MRR if relevant tags occur at early positions in the list of recommended tags.

**Mean average precision (MAP)** extends the precision metric and also considers the order of the recommended tags. This is done by computing the precision value at every position $k$ of the ranked list of tags and using the average of these values [20]:

$$MAP = \frac{1}{|U|} \sum_{u=1}^{|U|} \left( \frac{1}{|T_u|} \sum_{k=1}^{|t_u^k|} (B_k \cdot P@k) \right) \tag{13}$$

, where $B_k$ is 1 if the tag at position $k$ of the list of recommended tag is correct.

In particular, we report R@$k$, P@$k$, MRR and MAP for $k = 10$ and F1-Score ($F_1$@$k$) for $k = 5$ recommended tags.

### 5.3 Baseline Algorithms

We compared the results of our approach to several "baseline" tag recommender algorithms. The algorithms were selected in respect to their popularity in the community, performance and novelty [34]. The most basic approach we utilized is the unpersonalized *MostPopular (MP)* algorithm. MP recommends independent of user and resource, the same set of tags that is weighted by the frequency over all tag assignments [16]. A personalized extension of MP is the *MostPopular*$_{u,r}$ *(MP*$_{u,r}$*)* algorithm that suggests the most frequent tags in the tag assignments of the user (*MP*$_u$) and the resource (*MP*$_r$) [16]. As done in our approaches, we weighted the user and the resource components equally ($\beta = .5$).

Another well known recommender approach is *Collaborative Filtering (CF)* which was adapted for tag recommendations by Marinho et al. [34]. Here the neighborhood of a user is formed based on the tag assignments in the user profile and the only variable parameter is the number of users $k$ in this neighborhood. $k$ has been set to 20 as suggested by Gemmell et al. [30]. In Section 4.2 we have described how we applied *Latent Dirichlet Allocation (LDA)* for tag recommendations. The results presented in this work have been calculated using $Z = 1000$ latent topics [19].

An additional approach we utilized is the well-known *FolkRank (FR)* algorithm which is an improvement of the *Adapted PageRank (APR)* approach [16]. FR extends the PageRank algorithm in order to rank the nodes within the graph structure of a folksonomy [16], which is based on their importance in the network. Our implementation of APR and FR builds upon the code and the settings of the open-source Java tag recommender framework provided by the University of Kassel[11]. In this implementation the parameter $d$ is set to .7 and the maximum number of iterations $l$ is set to 10.

A different popular and recent tag recommender mechanism is *Pairwise Interaction Tensor Factorization (PITF)* proposed by Rendle & Schmidt-Thieme [18]. It is an extension of *Factorization Machines (FM)* and explicitly models

[11] http://www.kde.cs.uni-kassel.de/code

pairwise interactions between users, resources and tags. The FM and PITF results presented in this paper were calculated using the open-source C++ tag recommender framework provided by the University of Konstanz[12]. We set the dimensions of factorization $k_U$, $k_R$ and $k_T$ to 256 and the number of iterations $l$ to 50 as suggested in [18].

Finally, we tried to benchmark against two time-dependent approaches. The first one is the *GIRPTM* algorithm presented by Zhang et al. [8] which is based on the frequency and the temporal usage of a user's tag assignments. The approach models the temporal tag usage with an exponential distribution based on the first- and last-time usage of the tags. The second time-dependent tag-recommender approach is the *Base-Level Learning Equation with Context (BLL+C)* algorithm introduced in our previous work [9]. BLL+C is based on the ACT-R human memory theory by Anderson et al. [10] and uses a power-law distribution based on all tag usages to mimic the time-dependent forgetting in tag applications. In both approaches the resource component is modeled by a simple most popular tags by resource mechanism, as it is also done in our 3Layers approach. In previous work [9], we showed that BLL+C outperforms GIRPTM and other well-established algorithms, such as FR, PITF and CF.

The algorithms described in this section along with our developed approaches (see Section 4 are implemented within our Java-based *TagRec* framework [35]. Published as open-source software, it can be downloaded from our Github Repository[13] along with the herein used test and training sets (see Section 5.1 and 5.2).

## 6 Results

In this section we present the evaluation of the two novel algorithms in line with our research questions. In step 1, we compared the three 3Layers approaches (3L, $3LT_{topic}$ and $3LT_{tag}$) with one another, in order to examine our first research question of whether recency has a differential effect on topics and tags. According to the empirical analysis illustrated in Section 3, $3LT_{tag}$ yields more accurate predictions than $3LT_{topic}$ and 3L.

Results shown in Table 2 prove this assumption since - independent of the metric ($F_1$@5, MRR and MAP) and the number of LDA topics (100, 500, and 1000) applied - the difference between $3LT_{tag}$ and 3L is significantly larger than the one between $3LT_{topic}$ and 3L. This allows us to conclude that a user's gist traces (LDA topics) associated with the user's bookmarks are less prone to "forgetting" than a user's verbatim traces (tags associated with the bookmarks). Interestingly, this effect is more strongly pronounced under the narrow folksonomy condition (Flickr), where no tags of other users are available for the target user's resource, than under the broad folksonomy condition (BibSonomy and CiteULike), where users could get inspired by tags of other users.

Furthermore, Table 2 illustrates the performance of 3L, $3LT_{topic}$ and $3LT_{tag}$ for different numbers of LDA topics (100, 500 and 1000). It can be seen that

[12] http://www.informatik.uni-konstanz.de/rendle/software/tag-recommender/
[13] https://github.com/learning-layers/TagRec/

**Table 2.** $F_1$@5, MRR and MAP values for BibSonomy, CiteULike and Flickr showing the performance of 3L and its time-dependent extensions ($3LT_{topic}$ and $3LT_{tag}$) for 100, 500 and 1000 LDA topics (first research question).

| | # Topics | Measure | 3L | $3LT_{topic}$ | $3LT_{tag}$ |
|---|---|---|---|---|---|
| BibSonomy | 100 | $F_1$@5 | .197 | .198 | **.204** |
| | | MRR | .152 | .154 | **.161** |
| | | MAP | .201 | .202 | **.212** |
| | 500 | $F_1$@5 | .204 | .205 | **.209** |
| | | MRR | .156 | .158 | **.163** |
| | | MAP | .206 | .208 | **.215** |
| | 1000 | $F_1$@5 | .206 | .207 | **.211** |
| | | MRR | .157 | .158 | **.162** |
| | | MAP | .207 | .208 | **.214** |
| CiteULike | 100 | $F_1$@5 | .211 | .212 | **.221** |
| | | MRR | .192 | .194 | **.211** |
| | | MAP | .226 | .228 | **.248** |
| | 500 | $F_1$@5 | .218 | .219 | **.225** |
| | | MRR | .196 | .198 | **.211** |
| | | MAP | .232 | .234 | **.250** |
| | 1000 | $F_1$@5 | .232 | .233 | **.238** |
| | | MRR | .199 | .200 | **.212** |
| | | MAP | .235 | .236 | **.250** |
| Flickr | 100 | $F_1$@5 | .500 | .507 | **.535** |
| | | MRR | .421 | .429 | **.476** |
| | | MAP | .560 | .571 | **.634** |
| | 500 | $F_1$@5 | .564 | .567 | **.582** |
| | | MRR | .443 | .448 | **.476** |
| | | MAP | .591 | .596 | **.635** |
| | 1000 | $F_1$@5 | .568 | .571 | **.585** |
| | | MRR | .450 | .454 | **.477** |
| | | MAP | .599 | .604 | **.636** |

all three approaches provide good results for different levels of topic specialization, with the best accuracy values reached for 1000 LDA topics[14]. $F_1$@5, MRR and MAP values calculated for 1000 topics are further used within the second evaluation step, which is described in the next paragraph.

In a second step, we compared the performance of our approaches, especially $3LT_{tag}$, with several state-of-the-art algorithms. By this means we address our second research question, of whether 3L and its two extensions can be implemented in form of effective and efficient tag recommendation mechanisms. First, Table 3 reveals that all personalized recommendation mechanisms clearly outperform the unpersonalized MP approach. This is not surprising, as MP solely takes into account the tag's usage frequency independent of information about a particular user or resource.

Second and more important, 3L and its two extensions ($3LT_{topic}$ and $3LT_{tag}$) reach significantly higher accuracy estimates than the well-established mechanisms LDA, $MP_{u,r}$, CF, APR, FR, FM and PITF. From this we conclude that predicting tags based on psychologically plausible steps that turn a user's gist

[14] **NOTE:** We also performed experiments with more than 1000 LDA topics (e.g., 2000, 3000, ...). However, as also shown by related work (e.g., [19, 24, 25]) this step did not help in increasing the performance of the LDA-based tag recommenders.

**Table 3.** $F_1$@5, MRR and MAP values for all the users in the datasets (BibSonomy, CiteULike and Flickr) and for users with a minimum number of 20 bookmarks ($B_{min} = 20$) showing that our time-dependent 3LT$_{tag}$ approach outperforms current state-of-the art algorithms (second research question). The symbols $^{*}$, $^{**}$ and $^{***}$ indicate statistically significant differences based on a Wilcoxon Ranked Sum test between 3L, 3LT$_{topic}$, 3LT$_{tag}$ and BLL+C at $\alpha$ level .05, .01 and .001, respectively; $^{\circ}$, $^{\circ\circ}$ and $^{\circ\circ\circ}$ indicate statistically significant differences between our two time-dependent approaches 3LT$_{topic}$, 3LT$_{tag}$ and 3L at the same $\alpha$ levels.

| | $B_{min}$ | Measure | MP | LDA | MP$_u$ | MP$_r$ | MP$_{u,r}$ | CF | APR | FR | FM | PITF | GIRPTM | BLL+C | 3L | 3LT$_{topic}$ | 3LT$_{tag}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BibSonomy | - | $F_1$@5 | .013 | .097 | .152 | .074 | .192 | .166 | .175 | .171 | .122 | .139 | .197 | .201 | .206 | .207 | **.211** |
| | | MRR | .008 | .083 | .114 | .054 | .148 | .133 | .149 | .148 | .097 | .120 | .152 | .158 | .157 | .158 | **.162** |
| | | MAP | .009 | .101 | .148 | .070 | .194 | .173 | .193 | .194 | .120 | .150 | .200 | .207 | .207 | .208 | **.214** |
| | 20 | $F_1$@5 | .019 | .142 | .156 | .078 | .195 | .204 | .184 | .197 | .162 | .163 | .240 | .249 | .264 | .269 | **.296**$^{**}{}^{\circ}$ |
| | | MRR | .011 | .129 | .135 | .059 | .160 | .175 | .159 | .171 | .135 | .137 | .201 | .216 | .224 | .227 | **.251**$^{**}$ |
| | | MAP | .012 | .152 | .163 | .074 | .200 | .219 | .197 | .214 | .164 | .166 | .256 | .275 | .289 | .291 | **.325**$^{**}{}^{\circ}$ |
| CiteULike | - | $F_1$@5 | .007 | .068 | .182 | .033 | .199 | .157 | .162 | .160 | .113 | .130 | .207 | .215 | .232 | .233 | **.238**$^{**}$ |
| | | MRR | .005 | .065 | .164 | .024 | .179 | .168 | .181 | .181 | .116 | .149 | .196 | .205 | .199 | .200 | **.212** |
| | | MAP | .005 | .073 | .191 | .029 | .210 | .196 | .212 | .212 | .132 | .169 | .229 | .241 | .235 | .236 | **.250** |
| | 20 | $F_1$@5 | .008 | .145 | .228 | .031 | .237 | .228 | .221 | .225 | .193 | .196 | .282 | .298 | .331$^{*}$ | .334$^{*}$ | **.353**$^{***}{}^{\circ}$ |
| | | MRR | .006 | .144 | .225 | .022 | .233 | .271 | .237 | .239 | .201 | .210 | .321 | .335 | .312 | .316 | **.367**$^{**}{}^{\circ\circ\circ}$ |
| | | MAP | .006 | .162 | .258 | .028 | .269 | .308 | .273 | .276 | .229 | .237 | .369 | .389 | .369 | .373 | **.430**$^{**}{}^{\circ\circ\circ}$ |
| Flickr | - | $F_1$@5 | .023 | .169 | .435 | - | .435 | .417 | .328 | .334 | .297 | .316 | .509 | .523 | .568$^{***}$ | .571$^{***}$ | **.585**$^{***}{}^{\circ}$ |
| | | MRR | .023 | .171 | .360 | - | .360 | .436 | .352 | .355 | .300 | .333 | .445 | .466 | .450 | .454 | **.477**$^{*}{}^{\circ\circ\circ}$ |
| | | MAP | .023 | .205 | .468 | - | .468 | .581 | .453 | .459 | .384 | .426 | .590 | .619 | .599 | .604 | **.636**$^{*}{}^{\circ\circ\circ}$ |
| | 20 | $F_1$@5 | .030 | .190 | .382 | - | .382 | .495 | .322 | .334 | .309 | .309 | .534 | .553 | .610$^{***}$ | .616$^{***}$ | **.643**$^{***}{}^{\circ\circ\circ}$ |
| | | MRR | .028 | .174 | .322 | - | .322 | .473 | .309 | .317 | .290 | .289 | .485 | .508 | .478 | .485 | **.530**$^{**}{}^{\circ\circ\circ}$ |
| | | MAP | .029 | .215 | .427 | - | .427 | .655 | .405 | .419 | .378 | .376 | .664 | .701 | .661 | .670 | **.732**$^{***}{}^{\circ\circ\circ}$ |

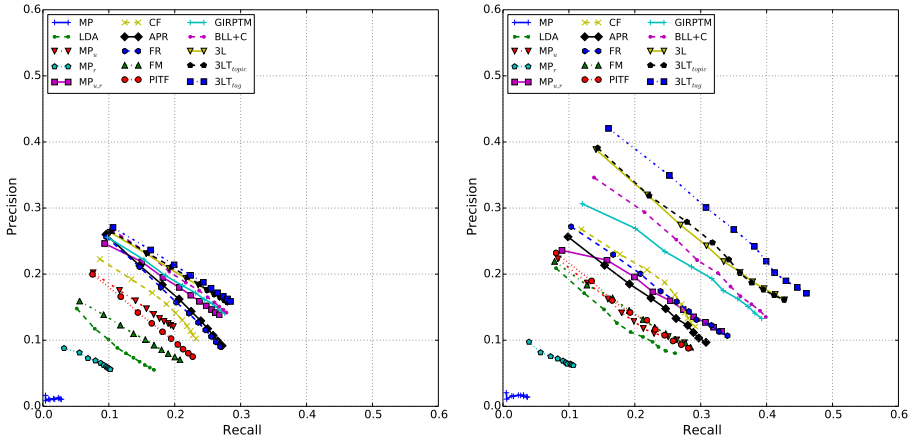
traces into words, calculates tag recommendations that correspond well to the user's tagging behavior.

Third, we can see that also the two other time-dependent algorithms (GIRPTM and BLL+C) outperform the state-of-the art approaches that do not take the time component into account. BLL+C based on ACT-R even reaches slightly higher estimates of accuracy than our 3L approach based on MINERVA2. However, this relation changes when we enhance 3L by the recency component at the level of tags. Then, $3LT_{tag}$ clearly outperforms BLL+C with respect to all three metrics and across all three datasets. Finally, as shown in Figure 3, a very similar pattern of results becomes apparent when evaluating the different approaches by plotting recall against precision for $k = 1$ - 10 recommended tags.

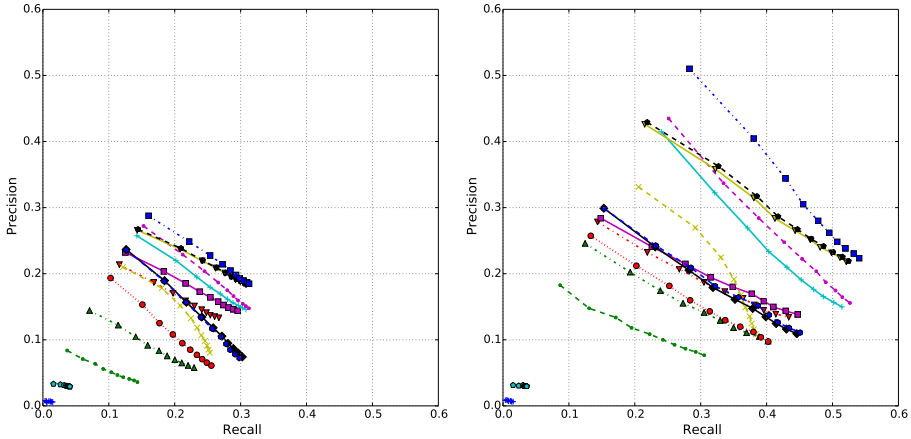
To furthermore prove our assumption that memory processes play an important role in social tagging systems, we also performed an experiment where we looked at users that have bookmarked a minimum of $B_{min} = 20$ resources (see also [36]). We conducted this experiment by applying a post-filtering method, i.e., recommendations were still calculated on the whole folksonomy graph but accuracy estimates were calculated only on the basis of the filtered user profiles (= 780 users in the case of BibSonomy, 1,757 in the case of CiteULike and 4,420 for Flickr). The results of the experiment are also shown in Table 3. We can observe that in general the accuracy estimates of all algorithms are increasing. Furthermore, it demonstrates that the difference between $3LT_{tag}$ and the other algorithms (including BLL+C) grows substantially larger the more user "memory" (history) is used. These differences between $3LT_{tag}$ and BLL+C as well as between $3LT_{tag}$ and 3L proved to be statistically significant based on a Wilcoxon Rank Sum test across all accuracy metrics ($F_1$@5, MRR and MAP) and all three datasets (see Table 3).

## 7 Discussion and Conclusion

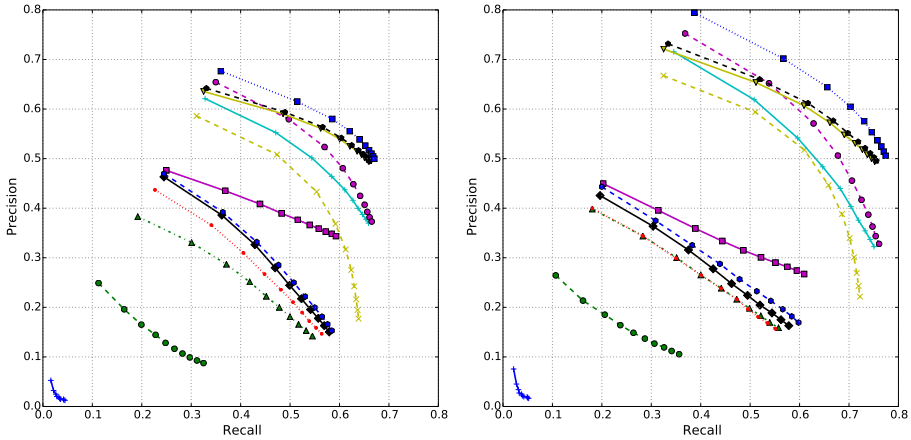
In this study we have provided empirical evidence for an interaction between the level of knowledge representation (semantic vs. lexical) and time-based forgetting in the context of social tagging. Based on the analysis of three large-scale tagging datasets (BibSonomy, CiteULike and Flickr), we conclude that - as expected - the gist traces of a user's personomy (the combination of LDA topics associated with the bookmarks) are more stable over time than the verbatim traces (the combination of associated tags). This pattern of results is well in accordance with research on human memory (e.g., [23]) suggesting that while people tend to forget surface details they keep quite robust memory traces of the general meaning underlying the experiences of the past (e.g., the meaning of read words). The interaction effect suggests that it is worthwhile to differentiate between both, time-based forgetting as well as the level of knowledge representation in social tagging research. Moreover, the differential affect of forgetting on the two levels of processing has further substantiated the differences between tagging behavior on a semantic level of gist traces and a lexical level of verbatim traces [28]. This in turn is in line with cognitive research on social tagging (e.g., [37]) that suggests




(a) BibSonomy (b) BibSonomy ($B_{min} = 20$)




(c) CiteULike (d) CiteULike ($B_{min} = 20$)




(e) Flickr (f) Flickr ($B_{min} = 20$)

**Fig. 3.** Recall/Precision plots for all the users in the datasets (BibSonomy, CiteULike and Flickr) and for users with a minimum number of 20 bookmarks ($B_{min} = 20$) showing the performance of the algorithms for 1 - 10 recommended tags ($k$).

to consider a latent, semantic level (e.g., modeled in form of LDA topics) when trying to understand the variance in the statistical patterns on the manifest level of users' tagging behavior.

Finally, we have gathered evidence for our assumption that interactive systems can be improved by basing them on a thorough understanding of how humans process information. We note in particular that integrating two fundamental principles of human information processing, time-based forgetting and differentiating into semantic and lexical processing, enhances the accuracy of tag predictions as compared to a situation when only one of the principles is considered. Our experiments showed that topics are more stable over time which means that they are, unless tags, not as suitable to be modelled using the BLL equation but can improve the results as an activation value on the basis of topic similarities. Therefore, 3L, that is based on the MINERVA2 theory of human categorization [1, 2] is enhanced by forgetting on the lexical level ($3LT_{tag}$). This approach significantly outperforms both the traditional 3L, as well as other well-established algorithms, such as CF, APR, FR, FM, PITF and the time-based GIRPTM. Furthermore, $3LT_{tag}$ also clearly reaches higher levels of accuracy than BLL+C, the to-date leading time-based tag recommender approach, that is based on the ACT-R theory of human memory [10] and was introduced in our previous work [9].

One limitation of this work is the calculation of semantic features (or topics) of the resources using LDA, which is not only very time-consuming but also could be biased because of the tag information it is based on. In this respect an interesting extension for future work would be to additionally conduct our experiments using external topics of the resources (e.g., Wikipedia categories as used in [5]). Looking at another aspect, our work has been inspired by the human memory model ACT-R proposed by Anderson et al. [10], but so far only investigates the first part of the equation, the recency component. Thus, it would be very interesting to further extend our approach by additionally investigating the associative component of the model. Also, as the computations were carried out with fixed values .5 for the exponent $d$ (3) and the weight $\beta$ (7), it would be worth exploring alternative values.

Moreover, we plan to include our algorithms in an actual online social tagging system (e.g., BibSonomy). Only in such a setting it is possible to test the recommendation performance by looking at user acceptance. Because our approach is theory-driven, it is rather straightforward to transfer it to recommendations in other interactive systems and Web paradigms where semantic and lexical processing plays a role (such as, for example, in Web curation). Thus, the generalization to other paradigms is another important benefit of driving recommender systems research by an understanding of human information processing on the Web.

**Acknowledgments** This work is supported by the Know-Center, the EU funded projects Learning Layers (Grant Agreement 318209) and weSPOT (Grant Agreement 318499) and the Austrian Science Fund (FWF): P 25593-G22. More-

over, parts of this work were carried out during the tenure of an ERCIM "Alain Bensoussan" fellowship programme.

## References

1. Hintzman, D.L.: Minerva 2: A simulation model of human memory. Behavior Research Methods, Instruments, & Computers **16** (1984) 96–101
2. Kwantes, P.J.: Using context to build semantics. Psychonomic Bulletin & Review **12** (2005) 703–710
3. Barsalou, L.: Situated simulation in the human conceptual system. Language and cognitive processes **18** (2003) 513–562
4. Glushko, R.J., Maglio, P.P., Matlock, T., Barsalou, L.W.: Categorization in the wild. Trends in cognitive sciences **12** (2008) 129–135
5. Seitlinger, P., Kowald, D., Trattner, C., Ley, T.: Recommending tags with a model of human categorization. In: Proc. CIKM '13, New York, NY, USA, ACM (2013) 2381–2386
6. Polyn, S.M., Norman, K.A., Kahana, M.J.: A context maintenance and retrieval model of organizational processes in free recall. Psychological review **116** (2009) 129
7. Anderson, J.R., Schooler, L.J.: Reflections of the environment in memory. Psychological Science **2** (1991) 396–408
8. Zhang, L., Tang, J., Zhang, M.: Integrating temporal usage pattern into personalized tag prediction. In: Web Technologies and Applications. Springer (2012) 354–365
9. Kowald, D., Seitlinger, P., Trattner, C., Ley, T.: Long time no see: The probability of reusing tags as a function of frequency and recency. In: Proc. WWW '14, New York, NY, USA, ACM (2014)
10. Anderson, J.R., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychological Review **111** (2004) 1036–1050
11. Helic, D., Trattner, C., Strohmaier, M., Andrews, K.: Are tag clouds useful for navigation? a network-theoretic analysis. International Journal of Social Computing and Cyber-Physical Systems **1** (2011) 33–55
12. Trattner, C., Lin, Y.l., Parra, D., Yue, Z., Real, W., Brusilovsky, P.: Evaluating tag-based information access in image collections. In: Proceedings of the 23rd ACM conference on Hypertext and social media, ACM (2012) 113–122
13. Körner, C., Benz, D., Hotho, A., Strohmaier, M., Stumme, G.: Stop thinking, start tagging: tag semantics emerge from collaborative verbosity. In: Proceedings of the 19th international conference on World wide web. WWW '10, New York, NY, USA, ACM (2010) 521–530
14. Lipczak, M.: Hybrid Tag Recommendation in Collaborative Tagging Systems. PhD thesis, Dalhousie University (2012)
15. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: Search and ranking. In: The semantic web: research and applications. Springer (2006) 411–426
16. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in folksonomies. In: Knowledge Discovery in Databases: PKDD 2007. Springer (2007) 506–514
17. Hamouda, S., Wanas, N.: Put-tag: personalized user-centric tag recommendation for social bookmarking systems. Social network analysis and mining **1** (2011) 377–385

18. Rendle, S., Schmidt-Thieme, L.: Pairwise interaction tensor factorization for personalized tag recommendation. In: Proc. WSDM 2010, New York, NY, USA, ACM (2010) 81–90
19. Krestel, R., Fankhauser, P., Nejdl, W.: Latent dirichlet allocation for tag recommendation. In: Proc. RecSys 2009, ACM (2009) 61–68
20. Rawashdeh, M., Kim, H.N., Alja'am, J.M., El Saddik, A.: Folksonomy link prediction based on a tripartite graph for tag recommendation. Journal of Intelligent Information Systems (2012) 1–19
21. Yin, D., Hong, L., Xue, Z., Davison, B.D.: Temporal dynamics of user interests in tagging systems. In: Twenty-Fifth AAAI conference on artificial intelligence. (2011)
22. Yin, D., Hong, L., Davison, B.D.: Exploiting session-like behaviors in tag prediction. In: Proc. WWW'2011, ACM (2011) 167–168
23. Brainerd, C., Reyna, V.: Recollective and nonrecollective recall. Journal of memory and language **63** (2010) 425–445
24. Kintsch, W., Mangalath, P.: The construction of meaning. Topics in Cognitive Science **3** (2011) 346–370
25. Krestel, R., Fankhauser, P.: Tag recommendation using probabilistic topic models. ECML PKDD Discovery Challenge 2009 (DC09) (2009) 131
26. Lorince, J., Todd, P.M.: Can simple social copying heuristics explain tag popularity in a collaborative tagging system? In: Proc. of WebSci '13, New York, NY, USA, ACM (2013) 215–224
27. Floeck, F., Putzke, J., Steinfels, S., Fischbach, K., Schoder, D.: Imitation and quality of tags in social bookmarking systems–collective intelligence leading to folksonomies. In: On collective intelligence. Springer (2011) 75–91
28. Seitlinger, P., Ley, T.: Implicit imitation in social tagging: familiarity and semantic reconstruction. In: Proc. CHI '12, New York, NY, USA, ACM (2012) 1631–1640
29. Helic, D., Körner, C., Granitzer, M., Strohmaier, M., Trattner, C.: Navigational efficiency of broad vs. narrow folksonomies. In: Proc. HT '12, New York, NY, USA, ACM (2012) 63–72
30. Gemmell, J., Schimoler, T., Ramezani, M., Christiansen, L., Mobasher, B.: Improving folkrank with item-based collaborative filtering. Recommender Systems & the Social Web (2009)
31. Doerfel, S., Jäschke, R.: An analysis of tag-recommender evaluation procedures. In: Proc. RecSys '13, New York, NY, USA, ACM (2013) 343–346
32. Campos, P.G., Díez, F., Cantador, I.: Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols. User Modeling and User-Adapted Interaction (2013) 1–53
33. Van Rijsbergen, C.J.: Foundation of evaluation. Journal of Documentation **30** (1974) 365–373
34. Balby Marinho, L., Hotho, A., Jschke, R., Nanopoulos, A., Rendle, S., Schmidt-Thieme, L., Stumme, G., Symeonidis, P.: Recommender Systems for Social Tagging Systems. SpringerBriefs in Electrical and Computer Engineering. Springer (2012)
35. Kowald, D., Lacic, E., Trattner, C.: Tagrec: Towards a standardized tag recommender benchmarking framework. In: Proc. HT'14, New York, NY, USA, ACM (2014)
36. Parra-Santander, D., Brusilovsky, P.: Improving collaborative filtering in social tagging systems for the recommendation of scientific articles. In: Proc. WI-IAT 2010. Volume 1., IEEE (2010) 136–142
37. Fu, W.T., Dong, W.: Collaborative indexing and knowledge exploration: A social learning model. IEEE Intelligent Systems **27** (2012) 39–46