

Generating Resource Profiles by Exploiting the Context of Social Annotations

Ricardo Kawase¹, George Papadakis^{1,2}, Fabian Abel³

¹ L3S Research Center, Leibniz University Hannover, Germany
kawase@L3S.de

² ICCS, National Technical University of Athens, Greece
gpapadis@mail.ntua.gr

³ Web Information Systems, TU Delft, The Netherlands
f.abel@tudelft.nl

Abstract. Typical tagging systems merely capture that part of the tagging interactions that enrich the semantics of tag assignments according to the system’s purposes. The common practice is to build tag-based resource or user profiles on the basis of statistics about tags, disregarding the additional evidence that pertain to the resource, the user or the tag assignment itself. Thus, the main bulk of this valuable information is ignored when generating user or resource profiles.

In this work, we formalize the notion of tag-based and context-based resource profiles and introduce a generic strategy for building such profiles by incorporating available context information from all parts involved in a tag assignment. Our method takes into account not only the contextual information attached to the tag, the user and the resource, but also the metadata attached to the tag assignment itself. We demonstrate and evaluate our approach on two different social tagging systems and analyze the impact of several context-based resource modeling strategies within the scope of tag recommendations. The outcomes of our study suggest a significant improvement over other methods typically employed for this task.

1 Introduction

One of the most popular innovations conveyed by Web 2.0 technologies is the introduction of *tagging*, a novel method of annotating resources with relevant keywords or terms in order to describe and enrich them with useful metadata. In *resource sharing systems* like Flickr¹, users mainly attach tags to their own resources, while *social tagging systems* like Delicious² enable users to create *tag assignments*³ for any resource shared with the community (i.e., free-for-all tagging [12]). Hence, there are two categories of tags: the *personalized* and the

¹ See <http://www.flickr.com>.

² See <http://www.delicious.com>.

³ A tag assignment is a user-tag-resource triple that describes which user assigned which tag to which resource.

collective ones [8]. Similarly, the benefits for the users are twofold: tags of the former category facilitate the organization and management of the resources, making search and retrieval more effective [11, 20]; collective tags, on the other hand, enhance the visibility of community content, associating relevant items with the same annotation(s) [8, 3].

Tag assignments are typically marked with subjectivity: different authors can interpret the same tag in different ways. Although this conveys significant benefits in the case of personalized tags, it also poses significant obstacles to the usefulness of the collective ones: the purpose of a tag assignment is not always clear to users other than its creator. For example, a tag associated with an image may describe it with respect to different aspects: the place and the persons depicted, the owner, an opinion or even its usage context (i.e., associated task). Thus, tags can be valid solely from a user-specific point of view [7]. This also explains why not all tags are suitable for search [4]; even those tags that mainly aim at describing the content of an item might characterize just a small part of the resource, without being representative of the entire resource. Some systems like LabelMe [17] and TagMe!⁴ [1] offer solutions to this problem by providing tags of finer granularity to their users.

In addition, tag assignments suffer from the ambiguity, inherent in any natural language: multiple meanings can be associated with the same tag (*polysemy*), while a specific tag can have multiple interpretations (*synonymy*). To disambiguate the meaning of tags for specific tag assignments, frameworks like MOAT [16] enable their users to associate each assignment with a URI specifying its meaning. This procedure is also incorporated in Faviki⁵, which uses Wikipedia as the source for URIs that clarify the meaning of an annotation (i.e., *semantic tagging*). A more flexible social tagging model is maintained in TagMe!, where users can enrich tag assignments with additional facets: semantic categories, URIs and spatial information. These facets represent contextual information that contribute to the disambiguation of the tag assignments, thus facilitating the search and the recommendation of tags or resources to a great extent.

In this paper, we argue that the aforementioned shortcomings of social annotations can be ameliorated by considering their context. In a previous work, we have already demonstrated the benefits of context for recommendation strategies [2]. However, the methods presented there were tailored to a particular system and, thus, were not generalizable to other social tagging systems. Instead, this work introduces a general, versatile modeling approach that builds comprehensive resource profiles, easily adapted to any folksonomy. It exploits the contextual information that is available in tagging systems rich in metadata, which are usually neglected.

At the core of our approach lies the idea of encapsulating not only the information that exclusively pertain to tags, but also additional contextual facets that refer to the other components of a tag assignment: the user, the resource and the

⁴ See <http://tagme.groupme.org>.

⁵ See <http://faviki.com>.

tag assignment itself. Merging these facets appropriately, we can derive weighted tag lists that form comprehensive contextual profiles, which are compatible and easily combined with typical tag-based profiles. These profiles can be employed in a diversity of common tasks that rely on tags, such as personalization, search and tag recommendation. We further describe how context-based profiles can be transformed into semantic URI-based profiles. We also put our generic resource modeling approaches into practice, demonstrating its applicability in two different social tagging systems: TagMe! and BibSonomy⁶. In both cases, we evaluate the impact of context-based profiles on the task of tag recommendations. The outcomes of our experimental study verify our premise that contextual profiles convey significant improvements in the performance of a social tagging system.

On the whole, the main contributions of our paper can be summarized as follows:

- we introduce the notion of tag-based and context-based resource profiles and present a generic context model for social tagging systems,
- we propose a generic strategy for exploiting context information embodied in social annotations, exemplifying it with a variety of resource modeling strategies, and
- we evaluate our strategies in two different tagging systems, verifying that the incorporation of contextual information clearly outperforms typical methods for generating resource profiles.

The remainder of the paper is structured as follows: in Section 2, we first elaborate on traditional folksonomies and recommenders proposed in the literature and then introduce our generic modeling strategy. Section 3 analyzes the potential benefits of exploiting context to model resources and describes the methodology of our experiments. In Section 4 we present and discuss the results of our evaluation, while in Section 5 we conclude the paper together with plans for future work.

2 Generating Resource Profiles

In the following, we elaborate on existing and novel strategies for generating generic resource profiles, which rely not only on the social annotations, but also on their context.

2.1 Modeling Social Annotations and Context

The structure that emerges from social annotations is called *folksonomy* [13]; it basically constitutes a set of user-tag-resource bindings, optionally coupled with the time each of them was performed [19]. In the context of our work, we consider the folksonomy model that was formally defined by Hotho et al. in [11].

⁶ See <http://www.bibsonomy.org>.

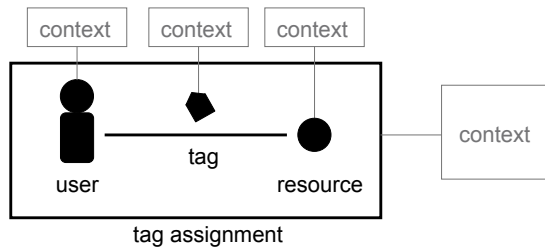


Fig. 1. Contextual information of social annotations can refer to the *user* that performed the tag assignment, to the *tag* that was designated by the user, to the *resource* that was annotated, or to the entire *tag assignment* itself.

Definition 1 (Folksonomy). A folksonomy is a quadruple $\mathbb{F} := (U, T, R, Y)$, where U, T, R are finite sets of instances of users, tags, and resources, respectively. Y defines the tag assignment, which is a relation between these sets (i.e., $Y \subseteq U \times T \times R$) that is potentially enriched with a timestamp indicating when it was performed.

The above definition abstracts from the tagging activities and does not incorporate contextual information. The latter refers either to the entities involved in a tag assignment (i.e., the user, the tag, and the resource), or to the tag assignment itself. This is clearly illustrated in Figure 1.

In the following, we consider all possible dimensions of contextual information: the meta-data attached to the tags, to the resources and to the users, as well as the usage context attached to tag assignments, as a whole. To cover the last case, we need to accommodate the attachment of any kind of context to a tag assignment. We employ an extension of Definition 1, namely the *context folksonomy model* [1].

Definition 2 (Context Folksonomy). A context folksonomy is a tuple $\mathbb{F} := (U, T, R, Y, C, Z)$, where:

- U, T, R, C are finite sets of instances of users, tags, resources, and context information, respectively,
- Y defines the tag assignment, which is a relation between U, T , and R (i.e., $Y \subseteq U \times T \times R$), and
- Z defines the context assignment, which is a relation between Y and C (i.e., $Z \subseteq Y \times C$).

2.2 Tag-based Profiles

At the core of this work lies the notion of folksonomy structures from the perspective of resources. Similar to a *personomy* (i.e., the user-specific part of a folksonomy, coined by Hotho et al. in [11]), we formally define the resource-specific fraction of a context folksonomy, called *personomy of a resource* from now on, as follows:

Definition 3 (Resource Personomy). *Personomy $\mathbb{P}_r = (U_r, T_r, Y_r, C_r, Z_r)$ of a given resource $r \in R$ is the restriction of \mathbb{F} to r , where U_r and T_r are the finite sets of users and tags, respectively, that are referenced from the tag assignments Y_r that are attached to r . C_r comprises the contextual information that are associated with the tag assignments in Y_r , and Z_r are the corresponding context assignments.*

In essence, a resource personomy encompasses the tag assignments that refer to a specific item along with their context. In a more abstract level, the *tag-based resource profile* $P(r)$ represents a resource as a set of weighted tags.

Definition 4 (Tag-based Resource Profile). *The tag-based profile $P(r)$ of a resource $r \in R$ is a set of weighted tags, where the weight of a tag t is computed by a certain strategy w with respect to the given resource r :*

$$P(r) = \{(t, w(r, t)) | t \in T, r \in R\}, \quad (1)$$

where $w(r, t)$ is the weight that is associated with tag t for a given resource r .

$P(r)@k$ denotes the subset of a tag-based profile $P(r)$ that contains the k tag-weight pairs with the highest weights. $\bar{P}(r)$ represents a tag-based profile whose weights are normalized, so that their sum is equal to 1, while $|P(r)|$ expresses the number of distinct tags contained in $P(r)$. It is worth clarifying at this point that the tags contained in $P(r)$ are not restricted to the tags that are explicitly associated with r (i.e., the tags included in the resource’s personomy \mathbb{P}_r). Instead, $P(r)$ may also specify the weight for a tag t_i that is associated to the resource r indirectly, through another element of its context. We illustrate this situation in Section 2.4 and Section 2.5 where we present our strategies for weighting tags.

In line with Definition 4, tag-based profiles can be built for a given user $u \in U$ and for a particular context $c \in C$, as well. For instance, tag-based user profiles (i.e., $P(u)$) have been studied by Firan et al. [6] and Michlmayr and Cayzer [14]. A straightforward approach to create a **tag-based context profile** $P(c)$ is to consider the tag assignments that pertain to c and to weight each of them according to the number of annotations that are contextualized with c and mention it. More formally: $w(c, t) = |\{(u, t, r) \in Y : (c, (u, t, r)) \in Z\}|$ (cf. Definition 2). In Section 2.5, we introduce more advanced strategies that exploit the characteristics of the respective type of context and show how these context profiles can be employed to enhance tag-based resource profiles.

2.3 Baseline Strategies for Tag-based Resource Profiles

The main challenge in generating tag-based profiles for resources is the definition of a strategy w that appropriately assigns weights to the involved tags. In the following, we present two weighting approaches that are typically used in the literature, but do not exploit all aspects of the context of tag assignments.

Tag Frequency. The rationale behind this approach is the assumption that the more users annotate a resource r with a tag t , the more salient is t for the description of r . Given the personomy of a resource \mathbb{P}_r , the corresponding tag-based resource profile $P(r)$ can be formed by counting the number of distinct users that assigned at least one tag $t \in T_r$ to the resource r . Hence, the weight $w(r, t)$ attached to a specific tag t in $P(r)$ is equal to: $w(r, t) = |\{u \in U_r : (u, t, r) \in Y_r\}|$. This approach was essentially employed by Cai and Li in [5] with the aim of improving tag-based personalized search.

Tag-based Co-Occurrence. In tagging systems like Flickr, resources are typically annotated with a limited number of distinct tags [18]. For this reason, Sigurbjörnsson and Zwol suggested in [18] to enrich the profile of a resource r with those tags that frequently co-occur with the tags assigned to r (i.e., T_r). The weight of those additional tags is equal to the frequency of their co-occurrence in the folksonomy:

$$w(r, t) = |\{(u, t_i, r_j) \in Y : \exists t_i \in T_r \wedge t \in T_{r_j}\}|.$$

The second method is typically employed in the context of tag recommendation techniques, which rely on association rules to capture the co-occurrence patterns (see, for instance, a recent, state-of-the-art method, introduced by Heymann et al. in [9]). For this reason, we employ it as the baseline method in our experimental study that examines the applicability of our algorithms in the tag recommendation task.

2.4 Generic Strategy for Generating Context-based Resource Profiles

Context-based resource profiling strategies rely on the contextual information available in folksonomies, and in resource personomies in particular: they build the profile of a resource r by merging (some of) the tag-based context profiles $P(c)$ associated with r . Moreover, one can also consider contextual information attached to the tag assignments referring to r (cf. Figure 1). The process of generating context-based resource profiles is outlined in the form of a generic approach in Definition 5.

Definition 5 (Context-based Resource Profile). *Given a tag-based profile $P(r)$ of a resource r and the set of tag-based context profiles $P(c_1), \dots, P(c_n)$, where $c_1, \dots, c_n \in C_r$ form the context information available in the resource personomy \mathbb{P}_r , the context-based resource profile $P_c(r)$ is computed by aggregating the tag-weight pairs (t_j, w_j) of the given profiles according to the following algorithm. Note that the parameter α_i allows for (de-)emphasizing the weights originating from profile $P(c_i)$.*

Input:

$P(r), ContextProfiles = \{(P(c_1), \alpha_1), \dots, (P(c_n), \alpha_n)\}$

Initialize: $P_c(r) = P(r)$

```

for  $(P(c_i), \alpha_i) \in \text{ContextProfiles}$ :
   $P(c_i) = \bar{P}(c_i)$ 
  for  $(t_j, w_j) \in P(c_i)$ :
    if  $(t_j, w_{P_c(r)}) \in P_c(r)$ :
      replace  $(t_j, w_{P_c(r)})$  in  $P_c(r)$  with  $(t_j, w_{P_c(r)} + \alpha_i \cdot w_j)$ 
    else:
      add  $(t_j, \alpha_i \cdot w_j)$  to  $P_c(r)$ 
    end
  end
end
Output:  $\bar{P}_c(r)$ 

```

The above algorithm is independent from the type of context information that is exploited to construct the context-based profiles and is, thus, generalizable to any tagging system. The construction of context-based resource profiles $P_c(r)$ depends, however, on the type of context that is considered. In the following, we present several weighting strategies for building them in systems rich in metadata, like TagMe! and BibSonomy.

2.5 Domain- and Application-specific Strategies for Generating Context-based Resource Profiles

TagMe! We begin with describing the strategies used to build contexts for resources in TagMe!. This system offers spatial tag assignments, enabling users to draw a rectangle that specifies the part of the image that is relevant to the corresponding tag. The resulting rectangular areas carry implicit information, which add more value to a tag assignment. Consider, for instance, the size and the distance of the tag’s area from the center of the resource; the former represents the portion of the visual space that is covered by the tag, with larger areas denoting tags that are more representative of the whole resource (i.e., tags with small area pertain to a particular object depicted in the picture, whereas large areas correspond to tags describing the picture in its entirety) [1]. Similarly, the latter expresses the relevance of tag assignments to the resource: tags closer to its center might be more important than tags placed at the margin of a resource [1].

In addition to this spatial facet, TagMe! provides two additional dimensions that are suitable for building context-based resource profiles: the categories and the semantic-meaning of tags. Categories can be freely entered by users via the tagging interface, in order to provide a more general description that disambiguates and describes tags more clearly. For instance, the tag “Brandenburger Tor” can be assigned to the category “Building”. In addition, TagMe! automatically enriches tags and categories assignments with DBpedia URIs to further disambiguate the meaning of a tag. In the following, we introduce strategies for building context-based profiles with the help of the tagging facets of TagMe!. Although the choice of these facets may seem rather intuitive, they have all been empirically evaluated in [1].

User-based Co-Occurrence. The rationale behind this weighting method is that an individual typically annotates similar resources, thus employing relevant tags in her tag assignments. This strategy considers, therefore, all users that assigned a tag to a given resource r and aggregates all the tags that they used (even for annotating other resources) into the context-based resource profile $P(r)$. The weight $w(r, t)$ is calculated by accumulating the frequencies of the tags available in the tag-based profiles of these users:

$$w(r, t) = \sum_{u \in U_r} |\{r_k \in R : (u, t, r_k) \in Y, r_k \neq r\}|.$$

Semantic Category Frequency. This strategy considers as evidence for the significance of a tag, the popularity of the category(ies) associated with the respective tag assignment(s). The premise here is that a tag associated with a category is more important than a tag without a category and, thus, more relevant to the annotated resource. In fact, the more frequent its category is, the more relevant it is. Thus, the weight of each tag is equal to the frequency of its category. In case a tag is associated with multiple categories, its weight amounts to the sum of the respective frequencies:

$$w(r, t) = \sum_i |\{(c_i, (u_j, t_k, r_l)) \in Z : \exists(c_i, (u, t, r)) \in Z_r\}|.$$

Co-occurring Semantic Category Frequency. The incentive for this strategy is the idea that tags described by the same categories are semantically relevant to each other. Consequently, when one of them is assigned to a particular resource r , the rest are also representative of r . Given a resource r , this weighting method retrieves all categories associated with r and places all tags associated with them (even through another resource) in the profile of r , $P(r)$. In line with the previous strategy, the value of each tag is set equal to the (sum of) frequency(ies) of the related category(ies):

$$w(r, t) = \sum_i |\{(c_i, (u_i, t_j, r_k)) \in Z : c_i \in C_r \wedge \exists(c_i, (u, t, r)) \in Z_r\}|.$$

Semantic Meaning. The rationale behind this approach is the assumption that semantically annotated tags constitute the more carefully selected annotations of a resource, thus being more representative of it and the basis for a more comprehensive description. Depending on whether a tag has been linked to a URI that uniquely identifies its meaning, this strategy defines two levels of importance. In other words, it assigns a binary value to each tag, with those tags that satisfy this condition receiving the value of 1, while the rest take the value of 0. More formally:

$$w(r, t) = 1 \text{ if } \exists(\text{URI}, (u, r, t)) \in Z_r.$$

Co-occurring Semantic Meaning. At the core of this strategy lies the idea that tags that are semantically equal to, but more popular than the tags directly associated with r , are more representative of its content. Thus, given a resource r , this strategy aggregates all the URIs involved in the tag assignments of r and builds the resource profile $P(r)$ by aggregating all tags that were associated with these URIs, independently of the respective resource. Tags are weighted according to the frequency(ies) of the URI(s) assigned to them: $w(r, t) = \sum_{\text{URI}_i \in C_r} |\{(\text{URI}_i, (u_j, r_k, t_l)) \in Z : \exists(\text{URI}_i, (u, r, t)) \in Z_r\}|.$

Area Size. The intuition behind this method is that the importance of tags is proportional to their size: the larger the area occupied by a tag, the more relevant the tag is to the annotated resource. On the other hand, tags that have been associated with a particular part of a resource, are considered more specific, and thus less significant. Thus, this strategy assigns to each tag a weight proportional to its area. More formally: $w(r, t) = |x_1 - x_2| \cdot |y_1 - y_2|$, where (x_1, y_1) and (x_2, y_2) are the Cartesian coordinates of the lower left and the upper right edge of the tag's rectangle ($x_1, x_2, y_1, y_2 \in [0, 1]$).

Distance From Center. This strategy is based on the assumption that the closer a tag is to the center of a resource (e.g., image), the more relevant it is. Hence, it weights tags according to their distance from the resource's central point, with smaller distances corresponding to higher values. Expressed mathematically, we have: $w(r, t) = \frac{1}{\sqrt{(x_{t_c} - x_{r_c})^2 + (y_{t_c} - y_{r_c})^2}}$, where (x_{r_c}, y_{r_c}) and (x_{t_c}, y_{t_c}) are the coordinates of the center of the resource and the center of the tag, respectively ($x_{r_c}, x_{t_c}, y_{r_c}, y_{t_c} \in [0, 1]$). Note that, with respect to the annotations of the above strategy, we have $x_{t_c} = \frac{x_1 + x_2}{2}$ and $y_{t_c} = \frac{y_1 + y_2}{2}$.

It should be stressed at this point that the above strategies rely on different facets of the context folksonomy of TagMe!. Thus, instead of being competitive to each other, they are complementary and can be arbitrarily combined. In total, we can have $(2^7 - 1) = 127$ distinct strategies, either *atomic* (i.e., composed of a single weighting method) or *composite* ones (i.e., derived from the combination of multiple weighting techniques).

BibSonomy. We now further demonstrate the adaptability and generality of our approach by proposing concrete context modeling strategies for the folksonomy of BibSonomy.

Co-occurring Journal Frequency. BibSonomy resources (i.e., publications) are typically associated with the journals or conferences, where they were published. This strategy exploits these metadata information, assuming that each specific journal is focused on a particular subject that represents the aggregation of similar resources. Thus, its publications are highly relevant to each other, and the tags assigned to one of them are probably applicable to the rest, as well. Given a resource r , this weighting method retrieves the *Journal* metadata associated with r and aggregates in $P(r)$ the tags of all the resources that were published by the same journal. The value of each tag is equal to its frequency:
 $w(r, t) = |\{(c_j, (u_j, t, r_i)) \in Z : \exists (c_j, (u, t, r)) \in Z_r\}|$, where c_j stands for the journal metadata of the given resource r .

Co-occurring Journal-Year Frequency. The rationale behind this strategy is the assumption that the topics of the papers published in a specific journal drift with the passage of time. As a result, the papers published in the same journal in a particular year are more relevant in with each other than with the papers published at a different point in time. In this context, this

weighting method retrieves for every resource r the *Journal* and *Year* meta-data associated with it; then, it generates a list of the tags of all resources that were also published within the same journal in the same year. Tag weights are set equal to the frequency of the tags:

$w(r, t) = |\{(c_{j,y}, (u_j, t, r_i)) \in Z : \exists (c_{j,y}, (u, t, r)) \in Z_r\}|$, where $c_{j,y}$ stands for the journal and year metadata of the given resource r .

2.6 Transforming Tag-based and Context-based Profiles into Semantic Profiles

The aforementioned context-based modeling strategies form the basis for the creation of **semantic profiles**; these are profiles that explicitly specify the semantics of a tag by means of URIs. For social tagging systems that assign meaningful URIs to tag assignments (e.g., TagMe!) or systems that make use of the MOAT framework [16] (e.g., LODr [15]), we propose the transformation of tag-based profiles into semantic profiles that, instead of a list of tags, consist of a weighted list of URIs.

It is worth noting at this point that the semantic meaning of tags depends on the context of their use. For example, the tag “paris” most likely refers to the city, but for some tag assignments it could also refer to a person. It is not possible, therefore, to have a global mapping of tags to URIs. Instead, it is necessary to map each individual tag assignment to a particular URI. Thus, we propose to transform the personomy \mathbb{P}_r (see Definition 3) and its tag assignments as follows:

Definition 6 (URI-based Resource Personomy). *Given the tag-based personomy $\mathbb{P}_r = (U_r, T_r, Y_r, C_r, Z_r)$ of a specific resource r and its URI assignments $Z_{r,uri} \subseteq Y \times C_{uri} \subseteq Z_r$, where C_{uri} is the set of URIs, the URI-based resource personomy, $\mathbb{P}_{r,uri} = (U_r, T_{r,uri}, Y_{r,uri}, C_r, Z_r)$, can be constructed by iterating over the tag assignments and replacing the tags with URIs of the corresponding URI assignments according to the following algorithm:*

```

 $T_{r,uri} = T_r \cup C_{uri}$ 
 $Y_{r,uri} = \{\}$ 
for  $(u, t, r) \in Y_r$ :
    for  $((u, t, r), uri) \in Z_{r,uri}$ :
         $Y_{r,uri} = Y_{r,uri} \cup (u, uri, r)$ 
    end
end
 $\mathbb{P}_{r,uri} = (U_r, T_{r,uri}, Y_{r,uri}, C_r, Z_r)$ 

```

Given the URI-based Resource Personomy and a URI-based Context Folksonomy (which can be constructed in a similar manner as the semantic personomy), we can apply the resource modeling strategies presented in Sections 2.3 and 2.5 in order to generate semantic resource profiles. In this way, the resource modeling framework presented above supports tag-based tasks in both the social tagging and the Semantic Web systems.

Tag Assignments (TAs)	1,288
TAs with Spatial Information	671
TAs with Category Information	917
TAs with URI Information	1,050
TAs with all information	432

Table 1. Technical characteristics of the TagMe! data set.

3 Experimental Setup

To measure the quality of the above, context-based resource modeling strategies, we apply them to the *tag recommendation* task: given a set of resources annotated with tags and metadata, the goal is to predict other tags that are also relevant to a specific resource, but have not yet been assigned to it. In the subsequent paragraphs, we describe the setup of the thorough, experimental evaluation we conducted in this context.

3.1 Social tagging data sets

In the course of our experiments, we employed two real-world data sets that stem from the aforementioned social tagging applications: TagMe! and BibSonomy. A detailed description of the technical characteristics of the data sets is presented below.

TagMe! This web application constitutes a multifaceted social tagging system that allows users to associate their annotations with a variety of (optional) metadata, which are suitable for building context-based resource profiles. The data we collected comprise the whole activity of the first three weeks after the launch of the system in June, 2009. In total, its user base comprises 30 users; half of them had a Flickr account and, thus, were able to tag their own pictures, while the rest assigned tags to random pictures and pictures of their own interest. A summary of the technical characteristics of this data set is presented in Table 1.

BibSonomy. BibSonomy [10] is a social bookmarking and publication-sharing system that has been running for over four years. The resources in Bibsonomy are publications, stored in BibTeX format. Each resource has several additional metadata, such as the corresponding journal, volume, year, as well as the author names. We employed Bibsonomy’s public data set that is available on-line from the 1st July 2010. It consists of 566,939 resources, described and annotated by 6,569 users. In total, there are 2,622,423 tag assignments and 189,664 unique tags. For our experimental study, we considered those resources that had the *journal* information and were tagged with at least five distinct tags. We randomly selected 500 of those resources and derived their context-based profiles from the entire data set.

3.2 Leave-one-out cross-validation

To evaluate the effect of context-based resource profiles on tag recommendations, we employed the leave-one-out cross-validation methodology in the following way: at each step, we hid one of the tag assignments and, then, we built the profile of the corresponding resource according to the selected strategy, based on the remaining assignments. The resulting profile encompasses a ranked list of tags, whose value is estimated according to the facets of the folksonomy that the current strategy considers. The goal is to predict the hidden tag by placing it in the top positions of the ranking.

To estimate the performance of the algorithms, we considered the following *metrics*:

Success Rate at 1 (S@1) denotes the percentage of tag predictions that had the missing tag at the first position of the ranking. It takes values in the interval $[0, 1]$, with higher values corresponding to higher performance.

Success Rate at 10 (S@10) stands for the percentage of tag predictions that had the missing tag in one of the top 10 positions of the ranking. Similar to $S@1$, it takes values in the interval $[0, 1]$, and the higher the value, the better the performance of the corresponding method.

As *baseline* strategies, we consider the approaches described in Section 2.3, which exclusively rely on the information encapsulated in tag assignments (i.e., user, tag, and resource). Note that the *tag frequency* strategy adds to a resource profile $P(r)$ only tags that have already been assigned to the resource. Consequently, it cannot be applied to the tag prediction problem without any further extension. Thus, we employ *tag-based co-occurrence* as the main baseline strategy and compare it to the context-based strategies of Section 2.5. These strategies enrich the traditional tag frequency with context-based profiles, following the process described in Definition 5.

4 Results

4.1 TagMe!

As mentioned above, the large number of facets of the TagMe! data leads to a total of 127 distinct context-based strategies. For the sake of readability and due to space limitations, we provide the results only for the atomic ones (see Definition 5) together with the best performing composite methods. It is worth noting at this point that our methods are employed as extensions to the baseline one, merging them with a weight α as described in Definition 5.

A summary of the performance of the baseline method and the atomic weighting strategies is presented in Table 2. It is evident that all context-based methods improve over the baseline, to a varying, but statistically significant extent ($p < 0.01$). The Semantic Description brings about a minor increase in $S@1$

Context ID	Context	S@1	S@10	Context Weight
0	Baseline	0.076	0.331	-
1	User-based Co-Occurrence	0.087	0.407	0.8
2	Spatial Annotation Size	0.089	0.408	0.4
3	Spatial Annotation Distance	0.094	0.377	0.9
4	Categorized Tag Frequency	0.085	0.352	0.5
5	Category-based Co-Occurrence	0.102	0.401	0.7
6	Semantic Description	0.078	0.407	0.8
7	Semantic-based Co-Occurrence	0.083	0.406	0.1

Table 2. S@1 and S@10 results for the atomic context-based strategies combined with the baseline in TagMe! data set.

of 2.6%, whereas the Spatial Annotation Distance and the Category-based Co-Occurrence account for an improvement well above 30%. Equally significant is the improvement with respect to the S@10 metric that varies from 6.3% for Categorized Tag Frequency context up to 23.3% for Spatial Annotation Distance. The latter indicates that annotations attached closer to the center of a resource are more valuable than those tags assigned to the margin (cf. Section 2.5).

The fourth column of Table 2 contains the optimal value of the weight used to merge the corresponding individual strategy with the baseline method. This value was determined through an exhaustive search of all values in the interval [0,2] with a step of 0.1. The actual effect of this parameter is demonstrated in Figure 2, where the performance for weight 1 (i.e., merging the baseline and the contextual strategy on an equal basis) is compared with the best performing weight. With the exception of the Semantic Description, we can notice that the

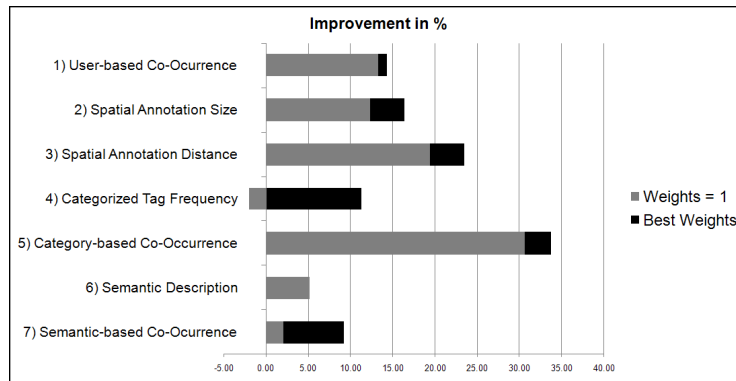


Fig. 2. S@1 improvement (in percentage) of each context over the baseline in the TagMe! data set. Gray bars show the results when the Context-Weight is set to 1, while black bars correspond to the performance of the best performing Context-Weight of each context.

ContextID (Context-Weight)	S@1	Improvement(%)
2(0.4) & 5(0.7) & 7(0.1)	0.106	38.8
2(0.4) & 3(0.1) & 5(0.7) & 7(0.1)	0.105	37.7
2(0.4) & 5(0.7)	0.105	37.7
5(0.7) & 7(0.1)	0.104	36.7

Table 3. S@1 results for the composite context-based methods that have the optimal performance on the TagMe! data set. ContextIDs refer to the methods of Table 2.

Context	S@1	Improvement(%)
Baseline	0.00712	-
Co-occurring Journal Frequency	0.00991	39.02
Co-occurring Journal-Year Frequency	0.01425	100.00

Context	S@10	Improvement(%)
Baseline	0.0701	-
Co-occurring Journal Frequency	0.0770	10.42
Co-occurring Journal-Year Frequency	0.1045	49.13

Table 4. S@1 and S@10 results for the baseline and the contextualized strategies (strategy-weight one to one) on the Bibsonomy data set.

calibration of this parameter conveys significant improvement, ranging from 2% for User-based Co-Occurrence to 12% for Categorized Tag Frequency.

Additionally, we experimented with all possible composite strategies (i.e., combinations of the atomic ones), employing again a variety of context-weights for each of them (i.e., $w_i \in [0, 2]$ for each method i). The best performing ones are presented in Table 3, along with the respective weight and the improvement they convey with respect to $S@1$. We can see that all of them perform significantly better than the individual methods comprising them. Note, though, that they all involve the atomic strategy with the highest value for $S@1$ (i.e., Category-based Co-Occurrence) and assign to it the highest weight. However, they improve its performance by just 2%. This clearly means that merging different contexts does not result in a cumulative improvement, because their combination leads to noise in the form of contradictory evidence: a tag rated high by a specific weighting strategy can be rated lower by another one.

In summary, we can conclude that contextualized strategies that rely on the spatial features, the categories and the semantics produce the best results in the case of TagMe!. They perform individually well enough and can be slightly improved when combined with the appropriate weights. Our semantic resource profiling strategies can be applied to any other semantic tagging system, thus being reusable and appropriate for other applications, as well.

4.2 BibSonomy

The use case of BibSonomy demonstrates how our model of context-based resource profiles can be beneficially applied to any folksonomy, and how we can

derive contextual information from the relations between the tag assignments. The outcomes of our evaluation are summarized in Table 4. We can see that both context-based methods substantially improve over the baseline, with the Co-occurring Journal-Year Frequency doubling its precision. Nevertheless, the overall success rate remains very low ($\sim 1\%$) in all cases. Note that the combination of the contextual weighting strategies with the baseline was done on an equal basis ($Context - Weight = 1$).

5 Conclusions

In this paper, we proposed novel approaches to generating and enriching resource profiles that exploit the multiple types of contextual information, available in most social tagging systems. We demonstrated that context can be derived from almost any metadata of the components of a tag assignment (i.e., user, tag, and resource) as well as from the tag assignment as a whole. We formalized the approach for modeling context-based profiles and described various, versatile strategies for combining them.

To verify the benefits of context-based resource profiles, we considered the task of tag recommendation, which typically relies on naive resource profiles that are derived from tag co-occurrences. We applied our strategies on two real-world datasets, with the outcomes indicating a considerable improvement over the baseline recommendation method. This verifies our premise that items sharing similar metadata (with respect to the same part of their tag assignments) are highly likely to be annotated with the same tags. We also demonstrated that contextual information pertaining to entire tag assignments provide significant evidence for modeling the resource profiles. This was proven to be particularly true for the cases where tag assignments are categorized, and spatially or semantically annotated.

Finally, we validated that merging different contexts does not result in a cumulative gain, since their arbitrary combination may lead to contradictory results. This issue actually lies at the core of our future work: we intend to develop techniques that identify complementary contexts, distinguishing them from the competitive ones. So far, there is no relevant work on this field, although techniques that optimally combine context models are expected to enhance the performance of many other tasks, as well, like personalization.

References

1. Abel, F., Henze, N., Kawase, R., Krause, D.: The impact of multifaceted tagging on learning tag relations and search. In: ESWC. pp. 90–105 (2010)
2. Abel, F., Henze, N., Krause, D.: Exploiting additional Context for Graph-based Tag Recommendations in Folksonomy Systems. In: WI 2008. pp. 148–154. IEEE (2008)
3. Ames, M., Naaman, M.: Why we tag: motivations for annotation in mobile and online media. In: Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '07). pp. 971–980. ACM, New York, NY, USA (2007)

4. Bischoff, K., Firan, C., Paiu, R., Nejd, W.: Can All Tags Be Used for Search? In: Proc. of Conf. on Information and Knowledge Management 2008. ACM (2008)
5. Cai, Y., Li, Q.: Personalized search by tag-based user profile and resource profile in collaborative tagging systems. In: CIKM. pp. 969–978 (2010)
6. Firan, C.S., Nejd, W., Paiu, R.: The benefit of using tag-based profiles. In: Almeida, V.A.F., Baeza-Yates, R.A. (eds.) LA-WEB. pp. 32–41. IEEE Computer Society (2007)
7. Golder, S., Huberman, B.A.: The structure of collaborative tagging systems. *Journal of Information Sciences* 32(2), 198–208 (2006)
8. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. *J. Inf. Sci.* 32(2), 198–208 (2006)
9. Heymann, P., Ramage, D., Garcia-Molina, H.: Social tag prediction. In: Myaeng, S.H., Oard, D.W., Sebastiani, F., Chua, T.S., Leong, M.K. (eds.) SIGIR. pp. 531–538. ACM (2008)
10. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: BibSonomy: A Social Bookmark and Publication Sharing System. In: Proc. First Conceptual Structures Tool Interoperability Workshop. pp. 87–102. Aalborg (2006)
11. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: Search and ranking. In: ESWC. Lecture Notes in Computer Science, vol. 4011, pp. 411–426. Springer (2006)
12. Marlow, C., Naaman, M., Boyd, D., Davis, M.: HT06, tagging paper, taxonomy, flickr, academic article, to read. In: Proc. of the 17th Conf. on Hypertext and Hypermedia. pp. 31–40. ACM Press (2006)
13. Vander Wal, T.: Folksonomy. <http://vanderwal.net/folksonomy.html> (2007)
14. Michlmayr, E., Cayzer, S.: Learning User Profiles from Tagging Data and Leveraging them for Personal(ized) Information Access. In: Proc. of the Workshop on Tagging and Metadata for Social Information Organization, 16th Int. World Wide Web Conference (WWW '07) (2007)
15. Passant, A.: LODr – A Linking Open Data Tagging System. In: First Workshop on Social Data on the Web (SDoW2008). vol. 405. CEUR Workshop Proceedings (2008)
16. Passant, A., Laublet, P.: Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data. In: Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008), Beijing, China (2008)
17. Russell, B.C., Torralba, A.B., Murphy, K.P., Freeman, W.T.: LabelMe: A Database and Web-based tool for Image Annotation. *International Journal of Computer Vision* 77(1-3), 157–173 (2008)
18. Sigurbjörnsson, B., van Zwol, R.: Flickr tag recommendation based on collective knowledge. In: Proc. of 17th Int. World Wide Web Conference (WWW '08). pp. 327–336. ACM Press (2008)
19. Wu, X., 0007, L.Z., Yu, Y.: Exploring social annotations for the semantic web. In: Carr, L., Roure, D.D., Iyengar, A., Goble, C.A., Dahlin, M. (eds.) WWW. pp. 417–426. ACM (2006)
20. Xu, S., Bao, S., Fei, B., Su, Z., Yu, Y.: Exploring folksonomy for personalized search. In: SIGIR 08. pp. 155–162. ACM, New York, NY, USA (2008)