Flexible Query Answering Systems: 8th International Conference, FQAS 2009, Roskilde, Denmark, October 26-28, 2009. Proceedings. Lecture Notes in Computer Science, Volumen 5822. Springer, 2009. 629-640.

DOI: http://dx.doi.org/10.1007/978-3-642-04957-6_54

Exploiting Social Tagging Profiles to Personalize Web Search

David Vallet, Iván Cantador, Joemon M. Jose

University of Glasgow, Glasgow, UK
{dvallet, cantador, jj}@dcs.gla.ac.uk

Abstract. In this paper, we investigate the exploitation of user profiles defined in social tagging services to personalize Web search. One of the key challenges of a personalization framework is the elicitation of user profiles able to represent user interests. We propose a personalization approach that exploits the tagging information of users within a social tagging service as a way of obtaining their interests. We evaluate this approach in Delicious, a social Web bookmarking service, and apply our personalization approach to a Web search system. Our evaluation results indicate a clear improvement of our approach over related state of the art personalization approaches.

1 Introduction

Nowadays, the size and pace of growth of information available to users constitute a difficult challenge for content retrieval technologies. The rapid propagation of the World Wide Web (WWW) has allowed users worldwide to have access to an unprecedented amount of information. Furthermore, in the WWW environment, there is a lack of strong global organization, with decentralized content provision, dynamic networks, etc., where query-based and browsing technologies often find their limits. Traditionally, users of Web search systems have described their information needs by providing a small set of keywords, with which the systems attempt to select the documents that best match these keywords. The majority of these queries are short (containing no more than 3 keywords on 85% of the times) and ambiguous [7], and often fail to represent the user's information need. Although the information contained in these keywords rarely suffices for the exact determination of user wishes, this is a simple way of interaction users are accustomed to; therefore, there is a need to investigate ways to enhance information retrieval, without altering the way they specify their requests. Consequently, information about user needs has to be found in other sources. It is in this scenario where personalized information retrieval can help the user to satisfy their information needs, using a range of personalization techniques that attempt to consider both the user's long and short term interests [4].

With the advent of the Web 2.0, social services have been exponentially increasing, in both terms of users and content. Some of these services allow users to provide annotations of resources. For instance, in Last.fm[1], users annotate their favourite songs; in Flickr[2], users store and tag their own photo streams; and in Delicious[3], users bookmark and tag interesting Web pages. Apart from facilitating the organization and sharing of content, these 'social tagging' actions can be a fairly accurate source user interests. Several studies have proven that a user profile can be effectively harvested from these tagging services [1,10], and later exploited on different personalization services, such as tag recommendation [3], item recommendation [9], and personalized search [6,9,12], to name a few.

In this work, we present a new personalized retrieval approach that makes use of a user profile defined within a social tagging service. The main research question investigated herein is whether Web search systems, such as Google or Yahoo!, can benefit from social tagging services. In particular, we investigate if a user profile defined in Delicious can be exploited to personalize a Web search system. Additionally, in order to evaluate our personalization approach, we propose an automatic technique to generate evaluation sets from social tagging corpora.

The rest of this paper is structured as follows. In Section 2, we define the Web search personalization model based on a social tagging profile, providing a brief comparison with the state of the art. In Section 3, we introduce our personalization approach. Section 4 describes our evaluation framework, and the followed experimental methodology. The results of our evaluation are presented in Section 5. Finally, Section 6 presents conclusions, together with possible future work.

2 Problem Definition

We first define a user and document profile model based on the underlying folksonomy of a social tagging service. A folksonomy F is defined as a tuple $\mathrm{F} = \{\mathrm{T}, \mathrm{U}, \mathrm{D}, \mathrm{A}\}$. $\mathrm{T} = \{t_1, \dots, t_L\}$ is the set of tags that comprise the vocabulary expressed by the folksonomy. $\mathrm{U} = \{u_1, \dots, u_M\}$ and $\mathrm{D} = \{d_1, \dots, d_N\}$ are respectively the set of users and the set of documents that annotate and are annotated with the tags of T. Finally, $\mathrm{A} = \{(u_m, t_l, d_n)\} \in \mathrm{U} \times \mathrm{T} \times \mathrm{D}$ is the set of assignments of each tag t_l to a document d_n by a user u_m. The profile of u_m is defined as a vector $\overrightarrow{u_m} = (u_{m,1}, \dots, u_{m,L})$ where $u_{m,l} = \left|\{(u_m, t_l, d\ ) \in \mathrm{A} | d \in \mathrm{D}\}\right|$ is the number of times the user has annotated resources with tag t_l. The profile of d_n is defined as a vector $\overrightarrow{d_n} = (d_{n,1}, \dots, d_{n,L})$ where $d_{n,l} = |\{(u, t_l, d_n\ ) \in \mathrm{A} | u \in \mathrm{U}\}|$ is the number of times the document has been annotated with tag t_l. In our Web search scenario, the set of documents D represents the resources present in the Web, and are identified by an URL. Users are identified by a user id.

We exploit the user and document profiles in order to personalize a Web search system. Let D be the set of documents present on the Web, a non-personalized Web search system S provides a rank list of documents $\mathrm{S}(q) \subseteq \mathrm{D}$ that satisfy a given query

[1] http://www.last.fm
[2] http://www.flickr.com
[3] http://www.delicious.com

topic q. The rank list will follow an ordering $\tau = [d_1 \geq d_2 \geq \cdots \geq d_k]$ where $d_i \in \mathrm{D}$, and $\geq$ is the ordering relation used by the search system. Similarly, we can define a personalization approach as a search system which provides a rank list of documents $\mathrm{S}(u) \subseteq \mathrm{D}$ that satisfy the preferences of user u. The personalization approach provides a final ordering $\tau' = [d_1 \geq d_2 \geq \cdots \geq d_k]$. The ordering relation is defined by $d_i \geq d_j \Leftrightarrow \mathrm{sim}(u, d_i) \geq \mathrm{sim}(u, d_j)$, where $\mathrm{sim}(u, d)$ is a similarity function between user u and document d. Typical personalization techniques implement this user-document similarity function.

2.1 Related Work

In this paper, we investigate whether a user profile defined in a social tagging service, such as Delicious, can be successfully exploited for personalized Web search. There have been previous studies which have investigated the use of the Delicious corpus in order to improve the retrieval process. For instance, Hotho et al. developed the FolkRank algorithm [4], an adaptation of the PageRank algorithm to the folksonomy structure. Among other applications, FolkRank proved to be a better popularity measure of a document than PageRank, as it exploits the user generated folksonomy, rather than the Web links. Bao et al. also investigated the use of a popularity measure derived from the folksonomy structure, but focused its application in a Web search system [2]. They introduced two importance score values, SocialSimRank and SocialPageRank, which calculate the relevance of a document to a query, and the popularity of a document, respectively. They concluded that these measures provide a better performance than traditional measures, such as term matching and PageRank. Similar to the studies of Hotho et al. and Bao et al., we exploit the folksonomy structure available on Delicious, but focus on offering a personalized search to the user, rather than improving the overall rank of documents. As Bao et al., we apply our approach to the Web search domain.

A personalized retrieval model that exploits user profiles defined in a folksonomy has been investigated in previous approaches [9,6,12]. Shepitsen et al. applied a hierarchical clustering algorithm to the tags associated to a user profile, defined in Delicious [9]. They used the generated tag clusters to provide personalized item recommendations. Rather than item recommendation, the approach presented in this paper follows a personalized retrieval model applicable to Web search, where a list of result are re-ranked according to the user preferences. This model is also followed by Xu et al. by presenting a user-document similarity function that relates the user and document tags [12]. Additionally, they presented a tag expansion approach, applied over a restricted corpus, which enriches the user profile representation. Noll and Meinel [6] also presented a personalized Web search model that exploits the user and document related tags, which improved a Web search system during their user evaluation. Our personalization approach follows the same personalization model as Xu et al.'s, and Null and Meinel's, but utilizes a different personalization technique to calculate the user-document similarity. Therefore, we compare and evaluate our proposed approach against the approaches presented by these authors.

3 Personalization Approaches Based on the Vector Spatial Model

This section presents the personalization approaches evaluated in our study. For convenience, Table 1 presents a number of definitions, which are standard weighting schemes used in the IR area and will be used by the presented personalization scores.

Table 1. Standard IR weighting schemes adapted to the folksonomy model

Description	Definition
User tag frequency	$tf_{u_m}(t_l) = u_{m,l}$
Document tag frequency	$tf_{d_n}(t_l) = d_{n,l}$
User-based tag inverse document frequency	$idf_u(t_l) = \log \frac{\mathrm{M}}{n_u(t_l)}, n_u(t_l) = \lvert\{u_m \in \mathrm{U} \vert u_{m,l} > 0\}\rvert$
Document-based tag inverse document frequency	$idf_d(t_l) = \log \frac{\mathrm{D}}{n_d(t_l)}, n_d(t_l) = \lvert\{d_n \in \mathrm{D} \vert d_{n,l} > 0\}\rvert$

We adopt the well known information retrieval Vector Space Model (VMS). The VSM represents user queries and documents as vectors in a finite space in order to calculate a similarity value between them. In Table 1, we define the tag frequency and inverse document frequency. These are an adaptation of the classic tf-idf weighting scheme, where the frequency of a term in the document (tf), and the inverse document frequency (idf) value of the term in the collection are considered. The term frequency follows the hypothesis that the more frequent a term is in a document, the more important this term is in describing the document. The inverse document frequency is a measure of the general importance of a term, meaning how common the term is in the collection of documents. In our model, we use the tag frequency instead of the term frequency.

Whereas the user and document tf define how important is the tag to the user and the document, respectively, we can disregard the document and user collections in order to calculate the global importance measure (such as idf). On the one hand, the user idf measure considers the importance of a tag by how common is the tag across users. On the other hand, the document idf measure considers the importance of a tag by how common is the tag across documents. Note that in the classic VSM, the document collection is the only source of the term's frequency and inverse document frequency. Analyzing our results, we will be able to conclude which of these measures is better to use on a personalization approach based on folksonomy user profiles. The approaches presented previously by Xu et al. [12] and Noll and Meinel [6] also follow the VSM. We present and evaluate their similarity functions, together with our own personalization technique.

3.1 Cosine Similarity Approach

The approach presented by Xu et al. use the classic cosine similarity measure to compute the similarity between user and document profiles. As weighting scheme, they use the $tf\text{-}idf$[4] value. Following our model, their approach can be defined as follows:

$$\cos_{tf\text{-}idf}(u_m, d_n) = \frac{\sum_l \left(tf_{u_m}(t_l) \cdot idf_{u_m}(t_l) \cdot tf_{d_n}(t_l) \cdot idf_{d_n}(t_l) \right)}{\sqrt{\sum_l (tf_{u_m}(t_l) \cdot idf_{u_m}(t_l))^2} \cdot \sqrt{\sum_l (tf_{d_n}(t_l) \cdot idf_{d_n}(t_l))^2}},$$

where the numerator is the dot product of the $tf\text{-}idf$ vectors associated to the user and the document, and the denominator is the user and document length normalization factors, calculated as the magnitude value of those vectors. Xu et al. compute a cosine similarity measure with a different weighting scheme, inspired by the BM25 retrieval model. We henceforth denote this measure as $\cos_{\mathrm{bm25}}(d_n, u_m)$. More information on this measure can be found in the authors' paper [12].

3.2 Scalar Tag Frequency Approach

The approach presented by Noll and Meinel differs from the previous in that it performs a scalar product eliminating the user and document length normalization factors [6]. Also, they do not make use of global tag importance measures, such as idf. They normalize all document tag frequencies to 1, since they state that the intention of this normalization is to give more importance to the user profile when computing the similarity measures, by only taking into consideration the matched tags between the user profile and the document associated tags. Following the notation given in Table 1, their similarity approach can be defined as follows:

$$tf(u_m, d_n) = \sum_{l: d_{n,l} > 0} tf_{u_m}(t_l) \ .$$

3.3 Scalar $tf\text{-}idf$ Approach

Next, we present our proposed personalization approach. Similarly to Xu et al.'s approach, we use the $tf\text{-}idf$ weighting scheme, but we eliminate the document and user length normalization factors. In the VSM, the finality of the length normalization factor is to penalize the score of documents that contain a high amount of information (i.e. a large quantity of terms). One of the drawbacks of this normalization factor is that short documents are usually ranked higher that larger ones, even if they have less terms in common with the user's query. In terms of tags, a document with a high number of related tags may mean that it is more popular for users, as more users have bookmarked it. Hence, if we used a length normalization factor, we would penalize the score of popular documents. In summary, whereas the document length in the

[4] Xu et al. do not specify if they use the user or document idf weights, or both. We chose to use both, as it gave the best performance values.

classic VSM denotes the amount of content presented, the document length in our model can be understood as a measure of the popularity of the document on the social tagging site. As several works point out, this popularity value is a good source of relevancy [2,4]. Thus, it would not be advisable to penalize popular documents. Note that eliminating the user length normalization factor does not have any effect, as it is constant in all user-document similarity calculations.

The main difference between our approach and Noll and Meinel's is that we incorporate the idf global tag importance factor, following the VSM idea that a more rare tag is more important when describing either the user's interests or the document's content. We neither normalize the content of the document, as we believe that the distribution of tags on a document may give insights on how important a tag is to describe the document's content. As mentioned previously, we can exploit two different sources in order to calculate the idf value associated to a tag: the user collection and the document collection. To investigate which is the best source for the idf measure, we present three variations of our approach:

$$tf\text{-}idf(u_m, d_n) = \sum_l \left(tf_{u_m}(t_l) \cdot idf_{u_m}(t_l) \cdot tf_{d_n}(t_l) \cdot idf_{d_n}(t_l) \right). \quad (1)$$

$$tf\text{-}idf_{u_m}(u_m, d_n) = \sum_l \left(tf_{u_m}(t_l) \cdot idf_{u_m}(t_l) \cdot tf_{d_n}(t_l) \cdot idf_{u_m}(t_l) \right). \quad (2)$$

$$tf\text{-}idf_{d_n}(u_m, d_n) = \sum_l \left(tf_{u_m}(t_l) \cdot idf_{d_n}(t_l) \cdot tf_{d_n}(t_l) \cdot idf_{d_n}(t_l) \right). \quad (3)$$

where Equation 1 makes use of the user idf measure on the user component and the document idf measure on the document component, Equation 2 uses the user idf measure on both components, and Equation 3 uses the document idf measure.

4 An Evaluation Framework for Personalized Web Search

Noll and Meinel [6] evaluated their personalization approach combined with a Web search engine. They adopted a user centred evaluation approach by creating a set of predefined queries, and by asking users to evaluate the results. More specifically, users were asked to evaluate which result list they preferred: either the Web search ranking or the personalized ranking. Xu et al. [12] used the social bookmarking information to create an automatic evaluation framework. The main advantage of their framework is that the experiments could be reproduced. However, they did not explore the performance of their personalization approaches when combined with a Web search engine. They combined their approach with a search system that was limited to the bookmarks pertinent to their test beds, ranging from 1K to 15K Web documents. The goal of our evaluation frameworks falls in the middle of these two approaches: 1) as Noll and Meinel, we are more interested in testing our personalization approach in a real Web search environment; and 2) as Xu et al., we adopt an automatic evaluation framework with a test bed of topics and relevance judgments extracted from the social bookmarking information. In this section, we describe our evaluation framework, highlighting the main differences between it and the previously presented evaluation frameworks.

4.1 Topic and Relevance Judgement generation

We split the tagging information of a given user into two parts. The first part forms the user profiling information, whereas the second is used for the automatic topic generation process. Hence, the subset of tag assignments used in the topic generation process is not included in the user profile, and thus are not part of our training data. This splitting process is applied to all users belonging to the initial test bed collection. Figure 1 outlines how the partition is made.

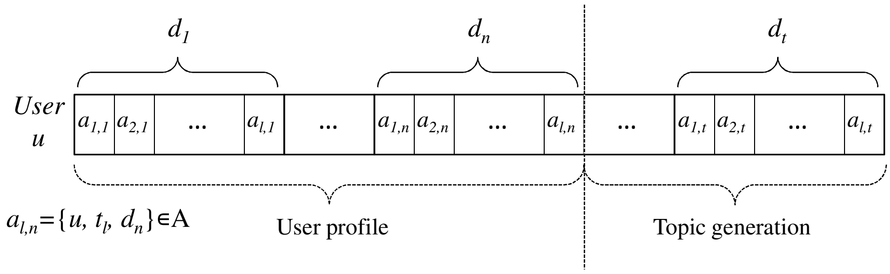

Fig. 1. Partitioning of user tag assignments into user profile and information intended for topic generation.

As shown in Figure 1, the topic creation process attempts to create a new topic from each annotated document $d \in [d_{n+1}, \ldots, d_t]$. The topic is defined by extracting the top most popular tags related to document d. We use the most popular tags as they are more objective to describe the document contents than those assigned by a single user. These tags are used to launch a Web search, and collect the result list obtained.

We then study how the different personalization approaches re-rank the returned result list. As document d was contained in the original user profile, we can assume that the document is relevant to the user. Thus, a good personalization approach will always rank the document in the top positions of the result list. We use the Mean Reciprocal Rank (MRR) [11] metric to measure the performance of our personalization approach. This measure assigns a value of performance for a topic of $1/r$, where r is the position of the relevant d in the final personalized result list. We also provide the P@N (Precision at position N) metric, which has a value of 1 iff $r \leq$ N. These values will be then averaged over all the generated topics. The topic generation and evaluation can be summarised in the following steps.

For each document $d \in [d_{n+1}, \ldots, d_t]$:

i. Generate a topic description using the top k most popular tags associated to the document.
ii. Execute the topic on a Web search system and return the top R documents as the topic's result list.
iii. If document d is not found in the result list, discard the topic for evaluation
iv. Apply the different personalization approaches to the result set.
v. Calculate MRR and P@N.

In our experiments, we used a query size of $k = 3$ tags, and a size result list of $\mathrm{R} = 300$ documents. Several studies point out an average user query size of 2-3 keywords in Web search [7]. We thus opted for a query size of three in order to emulate user using a Web search system, and to evaluate if user profiles obtained from the social tagging actions of the users could be successfully exploited to improve a Web search system. We also investigated the generation of topics with two keywords obtaining performance results similar to those obtained with topic sizes of three keywords. There is of course a chance that document d does not appear in the result list. In this case, the document is discarded for topic generation. With these settings, 23.8% of the topics were successfully generated, and the average position of document d on the result list was 65.4.

Xu et al. also built a test bed from the users' social tagging information. First, they applied the personalization techniques to a custom search engine that only retrieves documents that belong to the same test bed. On the other hand, we use a Web search system to return our documents, in this way we intend to have a more realistic set up. Second, they created the topic descriptions by using the tags associated to the user profile. They used a topic query size of one keyword (i.e. one tag belonging to the user profile), and made the assumption that if a document is tagged by the user with the same tag, it is relevant to the user. As we were using a Web search system to generate the topic results, using a single keyword very often failed to return any document that belonged to the user profile. This would have restricted our evaluation to documents that are highly popular (and thus are prone to be rank high on a single keyword query). We consider Xu et al.'s approach to be less restrictive than our approach; the topic definition are more broader, by only using one keyword to define them, and the relevance judgments are more loose by considering all documents tagged by the same tag. Our approach utilizes a more specific query, and restricts the relevance judgment to the document described by this query. Although we consider that our approach is more suited to evaluate a personalized Web search, both approaches could complement each other in order to give more insights on the performance of a personalization approach.

4.2 Experimental Setup

We create a test bed formed by 600 Delicious users. Delicious is a social bookmarking site for Web pages. As of the 26th of November of 2008, delicious had 5.3 million users[5], up from 1 million users registered on September of 2006[6]. With over 180 million unique URLs, delicious can be considered a fairly accurate "people's view" of the Web. This vast amount of user information has been previously successfully exploited in order to e.g. improve Web search [2], to provide personal recommendations [4,9], or to personalize search [6,12].

Due to limitations of Delicious API, we only extract the latest 100 bookmarks of each user, from which we use 90% of the bookmarks to create the user profile, and

[5] http://blog.delicious.com/blog/2008/11/delicious-is-5.html
[6] http://blog.delicious.com/blog/2006/09/million.html

the remaining 10% to generate the evaluation topics. The test bed contains 44,742 documents and 31,280 distinct tags. We did not apply any preprocessing steps to the user tags. Users used an average of 5.6 tags to describe each bookmark. As experimental Web search system, we use Yahoo!'s open Web search platform, Yahoo! Boss[7]. After the topic generation process, we ended up with 1,717 evaluation topics.

For each document in the topic result set, we downloaded the 100 most recent bookmarks. Those bookmarked documents had an average of 24.3 distinct associated tags. On average, 20.3% of the documents of the result list had been bookmarked at least once by a user. Figure 2 shows the distribution of this probability relative to the document position on the result list. Interestingly, this probability seems to stabilize at around 0.15 from the 200th position, which indicates that the proposed personalization approaches can be applied beyond the top results.

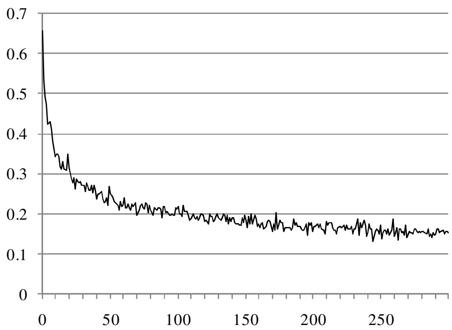

Fig. 2. Probability of a document being bookmarked relative to its position in the result list.

5 Experiment Results

In this section, we study the performance of the proposed personalization approaches within our evaluation framework. First, we make a comparison between all personalization approaches when applied to the evaluation topics. Second, we analyze the performance of these approaches when combined with the Web search results.

Table 2. Personalization approaches performance. Values with an asterix indicate a statistically significant higher value than the tf approach (Wilcoxon test, $p < 0.05$).

Metric	$\cos_{tf\text{-}idf}$	$\cos_{bm25}$	tf	$tf\text{-}idf_{d_n}$	$tf\text{-}idf$	$tf\text{-}idf_{u_m}$
MRR	0.0786	0.0991	0.2708	0.2878*	0.2989*	0.2990*
P@5	0.0897	0.1235	0.4281	0.4502*	0.4671*	0.4677*
P@10	0.1805	0.2155	0.6086	0.6290*	0.6325*	0.6302*
P@20	0.3512	0.3780	0.7734	0.7816	0.7880	0.7833

[7] http://developer.yahoo.com/search/boss/

Table 2 shows the MRR values and Precision at 5, 10 and 20 of the presented personalization approaches. Following the definitions of Section 2, this table shows the performance of the ranked result list according to the user's interests, namely $\mathrm{S}(u)$. The approaches are ordered in terms of the MRR metric. The performance of the approaches presented by Xu et al. [12], $\cos_{tf\text{-}idf}$ and $\cos_{\mathrm{bm25}}$, have much lower performance values than the rest of approaches, even though Xu et al. report a better performance of the tf approach, presented by Noll and Meinel [6]. The possible reason of this contradiction is that Xu et al.'s made use of controlled document collections, no larger than 15K documents, whereas in this evaluation and in the evaluation performed by Noll and Meinel's, a free Web search system was used to search for documents. As hypothesised in Section 3.3, the cosine similarity function penalizes those documents with a high amount of assigned tags (i.e. popular documents), in favour of documents in the result set that have fewer related tags. This penalization factor does not seem to work as good as the other approaches, which do not use the document length normalization factor.

All the variations of our personalization approach outperform the performance of Noll and Meinel's approach, tf. The improvement is statistically significant (Wilcoxon test, $p < 0.05$). We believe that this improvement is achieved thanks to the use of the idf value, which calculates the global importance of a tag. The obtained results are encouraging: we achieve a 10% improvement on the MRR metric with respect to the state of the art approach, tf, and a 9.2% improvement in terms of P@5. As explained in Section 3.3, the idf values can be computed from two sources, the user set or the document collection. By analyzing the results, we can conclude that the user set is a better source for the idf computation. The $tf\text{-}idf$ and $tf\text{-}idf_{u_m}$ personalization approaches use the user idf in the user component and both components, respectively. The difference on performance of these two approaches and the approach that use the document idf is also statistically significant.

We now investigate the performance of the personalization approaches when combined with the Web search results. In order to do this, we have to combine the result list returned by the Web search system with no personalization, denoted as $\mathrm{S}(q)$, with the result list produced by the personalization approaches, i.e. $\mathrm{S}(u)$. As a baseline, we use the Web search system results, but, in order to make a more fair comparison, we eliminated from the result list those documents that were not bookmarked by any user. The final ranked list is a combination of both the non-personalized and the personalized rank lists. We can define this combination as $\mathrm{S}(q,u) = \Psi\big(S(q), S(u)\big)$ where Ψ is a function that merges both ranked lists. We opted for a parameter free combination function, CombSUM, which is a rank based aggregation method [8].

Table 3 shows the performance values of the personalization approaches combined with the Web search. Values are correlated with those presented in Table 2. The cosine similarity personalization approaches degrade the performance of the Web search, while the rest of approaches perform better than the baseline. All the variations of the personalization approach introduced in this work outperform both the baseline and the best performing state of the art approach, i.e. tf. It is interesting to point that once our approach variations are combined with the

Web search results, there are no statistical differences between the performance of the approaches that use the user idf, and those exploit of the document idf. However, the $tf\text{-}idf_{u_m}$ approach, which uses only of the user idf, is the best performing, in terms of MRR. This approach achieves a 21.3% and a 4.5% improvement with respect to the baseline and tf approaches, respectively, with statistically significant differences. Again, the elimination of the document length normalization factor, and exploitation of the idf measure seems to be the key elements for these performance improvements.

Table 3. Personalization approaches performance when combined with the Web search engine result. Values with an asterix indicate a statistically significant higher value than the Web search ranking (Wilcoxon test, $p < 0.05$). Values marked with a † also indicate a statistically significant higher value than the tf approach.

Metric	baseline	$\cos_{tf\text{-}idf}$	$\cos_{\mathrm{bm25}}$	tf	$tf\text{-}idf_{d_n}$	$tf\text{-}idf$	$tf\text{-}idf_{u_m}$
MRR	0.3346	0.1573	0.1813	0.3885*	0.4023†	0.4026†	0.4060†
P@5	0.4607	0.2225	0.2638	0.5614*	0.5649*	0.5702*	0.5696*
P@10	0.5812	0.3809	0.4042	0.6832*	0.6820*	0.6907*	0.6913*
P@20	0.6948	0.5795	0.5649	0.7833*	0.7851*	0.7886*	0.7874*

6 Conclusions and Future Work

In this paper, we introduce an approach that exploits the user profile defined in a social tagging service to personalize a retrieval system. This personalization approach can be applied to any Web search system to provide personalization capabilities to any user who has a profile in a social tagging service, such as Delicious. This adds a new benefit of these services: with no extra effort, the user can take advantage of a personalized Web search system. In order to evaluate our approach, we propose an automatic test bed generation mechanism, which makes use of the tagging information available on the user profiles. The results of our evaluation our encouraging, and show that the adoption of global tag importance values, and the elimination of document length normalization factors significantly improves the state of the art personalization approaches, enhancing traditional Web search engines.

The popularity measure of a document is an important factor to measure its relevance to a query and a user. Although a Web search algorithm takes this importance factor into account (e.g. the PageRank measure), we should investigate how the folksonomy-based personalization approaches combine with folksonomy-based popularity measures, e.g. [4,2].

The folksonomy structure has been proven to be a good ground to expand the folksonomy-based user profiles [12], but these techniques are not scalable. A scalable expansion technique would allow its application to personalization approaches focused on Web search.

Acknowledgments. This research was supported by the European Commission under contract FP6-027122-SALERO. The expressed content is the view of the authors but not necessarily the view of the SALERO project as a whole.

References

1. Au-Yeung, C. M., Gibbins, N., Shadbolt, N.: A study of user profile generation from folksonomies. In: Proceedings of the WWW 2008 Social Web and Knowledge Management, Social Web Workshop (2008)
2. Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., Su, Z.: Optimizing web search using social annotations. In: Proceedings of the 16th international conference on World Wide Web (WWW 2007), pp. 501--510. ACM Press, New York (2007)
3. Chirita, P. A., Costache, S., Nejdl, W., Handschuh, S.: P-tag: large scale automatic generation of personalized annotation tags for the web. In: Proceedings of the 16th international conference on World Wide Web (WWW 2007), pp. 845--854. ACM, New York (2007)
4. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information Retrieval in Folksonomies: Search and Ranking. The Semantic Web: Research and Applications. LNCS, vol. 4011, pp. 411--426. Springer, Heidelberg (2006)
5. Micarelli, A., Gasparetti, F., Sciarrone, F., Gauch, S.: Personalized search on the world wide web. The Adaptive Web. LNCS, vol. 4321, pp. 195--230. Springer, Heidelberg (2007)
6. Noll, M. G., Meinel, C.: Web search personalization via social bookmarking and tagging. In: Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC 2007/ASWC 2007). LNCS, vol. 4825, pp. 367--380. Springer, Heidelberg (2007)
7. Jansen, B. J., Spink, A., Bateman, J., Saracevic, T.: Real life information retrieval: a study of user queries on the web. SIGIR Forum 32 (1), 5--17 (1998)
8. Renda, E. M., Straccia, U.: Web metasearch: rank vs. score based rank aggregation methods. In: Proceedings of the 2003 ACM symposium on Applied computing (SAC 2003), pp. 841--846, ACM Press, New York (2003)
9. Shepitsen, A., Gemmell, J., Mobasher, B., and Burke, R.: Personalized recommendation in social tagging systems using hierarchical clustering. In: Proceedings of the 2nd ACM Conference on Recommender Systems (RecSys 2008), pp. 259--266. ACM Press, New York (2008)
10. Szomszor, M., Alani, H., Cantador, I., O'hara, K., Shadbolt, N.: Semantic modelling of user interests based on cross-folksonomy analysis. In: Proceedings of the 8th International Semantic Web Conference. LNCS, vol. 5318, pp. pp. 632--648. Springer, Heidelberg (2008)
11. Voorhees, E.: The TREC-8 Question Answering Track Report. In: The Eighth Text REtrieval Conference (TREC 8), pp. 77--82 (1999)
12. Xu, S., Bao, S., Fei, B., Su, Z., Yu, Y.: Exploring folksonomy for personalized search. In: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR 2008), pp. 155--162, ACM Press, New York (2008)